University of Cyprus

DEPARTMENT OF ECONOMICS

Intergenerational Mobility: Econometric Theory and Applications

DOCTOR OF PHILOSOPHY DISSERTATION

ANTRI C. KONSTANTINIDI

2019

University of Cyprus

DEPARTMENT OF ECONOMICS

Intergenerational Mobility: Econometric Theory and Applications

ANTRI C. KONSTANTINIDI

A dissertation submitted to the University of Cyprus in partial fulfillment of the requirements for the degree of Doctor of Philosophy

May - 2019

VALIDATION PAGE

Doctoral Candidate: Antri C. Konstantinidi
Doctoral Thesis Title: Intergenerational Mobility: Econometric Theory and Applications

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Department of Economics and was approved on the 22.04.2019 by the members of the Examination Committee.

Examination Committee:

Research Supervisor: Andros Kourtellos, Assosiate Professor, University of Cyprus

Committee Member: Elena Andreou , Professor, University of Cyprus

Committee Member: Ioannis Bilias, Associate Professor, Athens University of Economics and Business

Committee Member: Steven Durlauf, Professor, University of Chicago

Committee Member: Dimitris Xefteris, Assistant Professor, University of Cyprus

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Antri C. Konstantinidi

Περίληψη

Ο στόχος αυτής της διατριβής είναι να μελετήσω την κοινωνική κινητικότητα χρησιμοποιώντας καινοτόμες οικονομετρικές μεθόδους που παρέχουν τόσο θεωρητική όσο και εμπειρική συμβολή στην υπάρχουσα βιβλιογραφία.

Στο Κεφάλαιο 1 μελετάται η μετάδοση της ευημερίας από γενιά σε γενιά, εστιάζοντας στον ρόλο των επενδύσεων των γονιών στα παιδιά κατά τη διάρκεια της παιδικής ηλικίας και της νεαρής ηλικίας, χρησιμοποιώντας δεδομένα από τη βάση δεδομένων *Panel Study of Income Dynamics* (*PSID*). Σε αντίθεση με την πλειονότητα της βιβλιογραφίας η οποία επικεντρώνεται σε έναν μόνο αριθμό που συνοψίζει την κινητικότητα μεταξύ γενεών, η ανάλυσή μας εστιάζει σε μια καμπύλη που καταγράφει τη διαγενειακή τροχιά κατά τη διάρκεια της πορείας ζωής ενός ατόμου. Χρησιμοποιούμε *functional data analysis*, μια μέθοδο ανάλυσης δεδομένων που μας επιτρέπει να κατασκευάσουμε εκτιμήσεις των τροχιών της κινητικότητας μεταξύ γενεών. Διαπιστώνουμε ότι οι γονικές επενδύσεις είναι πιο παραγωγικές στην πρώιμη και όψιμη εφηβική ηλικία ή στην νεαρή ενηλικίωση, ένα εύρημα το οποίο υποδεικνύει ότι η χρονική στιγμή κατά την οποία γίνονται επενδύσεις στην εκπαίδευση και το ανθρώπινο κεφάλαιο είναι πολύ σημαντική. Επιπλέον, υπάρχουν στοιχεία ετερογένειας λόγω της κοινωνικοοικονομικής κατάστασης και της οικογενειακής δομής των παιδιών. Τέλος, ο χρόνος κατά τον οποίο τα παιδιά που προέρχονται από δυσπραγούσες οικογένειες βιώνουν κάποιο σοκ (διαζύγιο, θάνατος ενός γονιού ή οικονομικό σοκ) αποτελεί σημαντικό παράγοντα για την ανοδική τους κινητικότητα.

Στο Κεφάλαιο 2 αναπτύσσουμε μια νέα τάξη μοντέλων κοινωνικής αλληλεπίδρασης που γενικεύουν το *Spatial autoregressive model* έτσι ώστε να επιτρέπει την παρουσία ετερογένειας με τη μορφή *threshold effects*. Αυτά τα μοντέλα μπορούν να εφαρμοστούν για να εξηγήσουν μια σειρά μη γραμμικών φαινομένων όπως περιπτώσεις όπου τα άτομα παραμένουν παγιδευμένα στη φτώχεια (*poverty traps*). Συγκεκριμένα, προτείνουμε ένα γενικό μοντέλο *Threshold Spatial Autoregressive* (*TSAR*), το οποίο αποτελεί γενίκευση τόσο του *Mixed regressive, spatial autoregressive model* όσο και του *Spatial autoregressive model* και επιτρέπει την ύπαρξη ενδογενών κοινωνικών αλληλεπιδράσεων ανά διαφορετικό καθεστώς. Αναπτύσσουμε μια μέθοδο *GMM* σε δύο βήματα για την εκτίμηση των παραμέτρων του μοντέλου και δείχνουμε την συνέπεια και την ασυμπτωτική κανονικότητα των προτεινόμενων εκτιμητών. Τέλος, αξιολογούμε την απόδοση των μεθόδων μας χρησιμοποιώντας *Monte Carlo* προσομοιώσεις.

Στο Κεφάλαιο 3 μελετάμε τη στατιστική συμπερασματολογία σχετικά με τις *threshold* παλινδρομήσεις ενώ υπάρχει αβεβαιότητα ως προς ποιο είναι το πραγματικό μοντέλο. Αυτό το πρόβλημα προκύπτει όταν κάποιος ενδιαφέρεται να ελέγξει την ύπαρξη μη γραμμικοτήτων τύπου *threshold*, αλλά υπάρχει αβεβαιότητα σχετικά με το σύνολο των μεταβλητών που πρέπει να συμπεριληφθούν στο μοντέλο. Η τυπική προσέγγιση για την αντιμετώπι-

ση της αβεβαιότητας του μοντέλου είναι η $post - single$ προσέγγιση, δηλαδή αρχικά η επιλογή των μεταβλητών ελέγχου (για παράδειγμα μέσω κάποιων ελέγχων υποθέσων) και στη συνέχεια η εξαγωγή συμπερασμάτων. Εντούτοις, η $post - single$ προσέγγιση οδηγεί σε σοβαρές στρεβλώσεις στο μέγεθος και στην ισχύ ενός ελέγχου για ύπαρξη μη γραμμικοτήτων τύπου *threshold*. Στο παρόν κεφάλαιο υιοθετούμε την $post - double$ προσέγγιση των *Belloni*, *Chernozhukov*, και *Hansen* (2011) στο πλαίσιο των *threshold* παλινδρομήσεων και δείχνουμε ότι ο ελεγχος υποθέσων μετά τη χρήση αυτής της μεθόδου λειτουργεί καλά τόσο σε μέγεθος όσο και σε ισχύ. Τέλος, αυτό το κεφάλαιο αξιολογεί την απόδοση της προτεινόμενης μεθόδου μέσω προσομοίωσης *Monte Carlo*.

Abstract

The broad aim of my thesis is to study social mobility using novel econometric methods that provide both theoretical and empirical contributions in the existing literature.

In Chapter 1 we study the intergenerational transmission of well-being by focusing on the role of trajectories of exposures during childhood and young adulthood using PSID data. Our analysis shifts the focus from a single number that summarizes the intergenerational mobility to a curve that captures the intergenerational trajectory over the life-course of an individual. In doing so, we employ a functional data analysis approach that allows us to construct estimates of trajectories of intergenerational mobility. We find that parental investments are more productive in the early and late childhood or young adulthood, highlighting the importance of the timing of human capital investments. Furthermore, we uncover evidence of heterogeneity due to socioeconomic status and family structure that suggests that the timing of the shocks for the disadvantaged children is an important factor for their upward mobility.

In Chapter 2 we develop a new class of social interaction models that generalize the spatial autoregressive model to allow for threshold effects. These models can be applied to explain a range of nonlinear phenomena such as poverty traps. In particular, we propose a general Threshold Spatial Autoregressive (TSAR) Model, which nests both mixed regressive, spatial autoregressive model as well as the spatial autoregressive model and allows for regime specific endogenous as well as contextual effects. We develop a two-step GMM method for the estimation of the threshold and regression parameters and show consistency and asymptotic normality of the proposed estimators. Finally, we assess the performance of our methods using a Monte Carlo simulation.

In Chapter 3 we study inference in threshold regressions in the presence of model uncertainty. This problem arises when one is interested in testing for the presence of threshold type nonlinearities but there exists uncertainty about the set of controls. The standard approach to deal with model uncertainty is the post-single approach, that is, select the control variables and then draw an inference. However, post-single selection leads to severe size and power distortions. Following Belloni, Chernozhukov, and Hansen (2011) this chapter uses a post-double selection procedure to construct a threshold test that is valid under model uncertainty and performs well in both size and power. Finally, this chapter evaluates the finite sample performance of the proposed method via a Monte Carlo simulation.

Acknowledgements

I would like to express my gratitude to many people who along with their advice, time and knowledge, have helped me complete this long journey.

Firstly, I would like to express my deepest gratitude to my advisor Andros Kourtellos for the continuous support of my Ph.D study, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I would like to express my gratitude to the members of the examination committee, Elena Andreou, Ioannis Bilias, Steven Durlauf, and Dimitris Xefteris for their time, comments and suggestions.

Furthermore, I gratefully acknowledge Yiguo Sun for her guidance, consistent involvement and valuable comments.

My deepest thanks also goes to Snezana Eminidou, Almarina Gramozi and Kyriakos Petrou for the constant support, for the days we were working together, and for all the good times we have had in the last four years. Special thanks to Kyriakos Petrou for his valuable feedback and the time we spent together discussing ideas.

I would also like to express my graditude to the whole academic and administrative staff of the department, for the support they have offered me. Special thanks to Anastasia Demetriou, and Litsia Tsiali.

Last but not least, I would like to thank my family, because without their endless love and selfless support, I would not be able to accomplish anything. My father Christos, my mother Astero, and Sofia.

To my parents and Sofia

Contents

List of Figures

List of Tables

Introduction

Social mobility is one of the most important aspects of inequality and refers to how the socio-economic position (outcomes and socioeconomic characteristics) of the children as they grow-up to become adults relates to that of their parents. While social mobility captures both the intragenerational and intergenerational aspects of the transmission of socioeconomic status, the latter is of particular importance as it can be used to study poverty traps. Despite the vast work, the literature has not provided satisfactory answers to the questions of persistent inequality and poverty traps. My thesis addresses various challenges by proposing three complementary approaches that focus on intergenerational trajectories of income, threshold-type nonlinearities due to neighborhood effects and model uncertainty in the context of threshold regressions.

Specifically, in the first chapter we study the intergenerational trajectories of the offsprings. Chapter 2 develops an econometric model which allow us to study poverty traps in the context of social networks and provides an empirical illustration. Chapter 3 studies inference on threshold regressions under model uncertainty. Chapter 2 has given rise to a joint paper with Andros Kourtellos and Yiguo Sun while Chapters 1 and 3 have generated a number of joint papers with Andros Kourtellos.

In Chapter 1, we shift the focus from a single number that summarizes the intergenerational mobility to a curve that captures the intergenerational trajectory over the life-course of an individual using functional data analysis. The standard empirical approach in the economics literature on intergenerational mobility focuses on intergenerational elasticity of income (IGE). The IGE is the slope of the coefficient of a log-log linear regression model of child's permanent income on parent's permanent income controlling for some characteristics. Its magnitude determines the degree of intergenerational mobility; for example, a value close to 1 implies greater persistence of the intergenerational transmission of income, suggesting lower mobility. However, there is no apriori reason to believe that a simple average of observed income is a sufficient statistic for permanent income since it ignores important mechanisms that affect offspring's income. Several studies have investigated the importance of critical periods in the human capital development of a child (Cunha and Heckman (2007) and Cunha, Heckman, and Schennach (2010)), as well as the dynamic complementarity in investments in different periods and the interaction with the timing of borrowing constraints (Caucutt and Lochner (2017)). In this chapter, we examine the role of the timing of parental income during childhood and young adulthood years using annual or biannual data from the Panel Study of Income Dynamics. Additionally, we investigate the heterogeneity in the intergenerational trajectories of income due to the socioeconomic status and structure of the family. We propose the intergenerational trajectories model using functional regression and we estimate an intergenerational elasticity function, which captures the intergenerational effect of the resources available to child at age t. Overall, we find that the parental income in early and late childhood is important for the outcomes of children as adults, while parental investments for young adulthood can be at least as productive as the ones in early and late childhood. The socioeconomic background of the parents affects the intergenerational trajectories of the offspring and the timing of parental income shocks plays a key role in offsprings' long run outcomes. Finally, early family shocks seem to affect more parental investments, making them less productive. This chapter contributes to the literature of intergenerational mobility by providing a complementary approach that focuses on intergenerational trajectories of income, taking into account for heterogeneity with respect to socioeconomic status of parents and the effects of income and family 'shocks'.

In Chapter 2, motivated by our empirical work on intergenerational mobility, we propose a new class of social interaction models that generalize the Spatial Autoregression-Mixed Regression to allow for threshold effects that capture the heterogeneity in the endogenous social interaction effects. Threshold-type nonlinearity is suggested by several economic theories such as models of income dynamics and poverty traps and more generally by models that feature multiple equilibria due to incomplete markets, increasing returns, complementarities, etc. One particular class of models that produces threshold-like nonlinearities are models of neighborhood effects. For example, threshold-like behavior can arise in a model with strict stratification of neighborhoods by income (Benabou (1996) and Durlauf (1996a,b)). We consider a two-step GMM estimator and develop an asymptotic distribution theory for the GMM estimators of the regression parameters as well as for the threshold parameter. While the first step GMM estimator is consistent, it is not efficient. Hence, we propose a second step estimator that aims at addressing this issue by obtaining a quasi-optimal estimator. Our framework nests both the fixed and diminishing threshold effect, and the threshold parameter estimate is normally distributed. We provide Monte Carlo simulations that show the finite sample performance of our estimators. Moreover, we propose a test for the threshold effect which features the Davies problem (that the threshold parameter is not identified under the null). The econometric methodology developed in this chapter is applicable to many interesting phenomena such as poverty traps. This chapter contributes to the literature of Spatial Autoregression-Mixed Regression by considering threshold-type nonlinearities.

Chapter 3 addresses the issue of model uncertainty with respect to the variables selection in the context of threshold regressions. In many economic contexts, applied economists have little guidance on the variables that should be included in the model, mainly due to the fact that the validity of one theory does not logically exclude other mechanisms from also being relevant. The usual approach is to select variables using a test for the statistical significance or selection criteria or a shrinkage and selection method in high dimensional framework and then conduct inference as if the real model was chosen. However, the choice of the variables affects the inference about the threshold. We are extending the post-double selection method of Belloni, Chernozhukov, and Hansen (2011), to the threshold regressions framework. The ideas of post-double selection method is based on the partialling out technique of Frisch-Waugh-Lovell in the linear setting, the Neyman's $C(\alpha)$ test in the nonlinear setting (Neyman (1979)), and Robinson (1988) in the semi-parametric setting. First, it is shown that the standard post-single selection methods have adverse effects in the size and power of the bootstrap-based threshold test proposed by Hansen (1996). Then, we proceed in showing how the proposed post-double selection procedure restores the distortion of the size and power of the threshold test. Monte Carlo simulations suggest that post-double selection restores the size and the power of the relevant bootstrap threshold test. This chapter contributes to the literature of the threshold regressions framework by proposing a post-double selection procedure to construct a threshold test that is valid under model uncertainty. This methodology can be applied to a range of interesting applications, in both empirical microeconomics and macroeconomics (e.g., intergenerational mobility, child development literature, and cross-country growth studies).

Chapter 1

Intergenerational Trajectories

1.1 Introduction

The standard empirical approach in the economics literature on intergenerational mobility focuses on intergenerational elasticity of income (IGE) which is the slope of the coefficient of a linear regression model of child's permanent income on parent's permanent income controlling for some characteristics. The magnitude of the IGE coefficient determines the degree of intergenerational mobility. An IGE close to zero implies greater mobility while an IGE value close to one implies higher degree of persistence, that is, immobility across generations. While the IGE model is a statistical model, under certain assumptions the linear IGE model can be interpreted as a behavioral model, which is implied by the classical theory of family income or investment models; see, for example, Becker and Tomes (1979). Broadly speaking these models focus on the intergenerational transmission mechanism of income, which stems from the dependence of education investments in parental income. This model is often augmented with credit constraints that capture the idea that parents cannot borrow against children's future income because they cannot provide credible repayment assurances (Becker and Tomes (1986) and Loury (1981)).

One key aspect in the implementation of the IGE model is the measurement of permanent income which is latent. The standard empirical practice measures permanent income as an average over several years. The idea is that transitory components average out to zero if a large enough time horizon is used. However, there is no reason to assume that transitory components average out to zero over the lifetime of an individual. For example, if the variance of the transitory component is not constant over the life cycle of the individual, then the age at which father's earnings are measured is also important.[1] More importantly, using a simple average as a proxy of permanent income completely ignores the importance of critical and sensitive periods in the development of a child. Cunha and Heckman (2007) and Cunha, Heckman, and Schennach (2010)) emphasize the importance of the timing of human capital investments and dynamic complementarity in investments. In fact the dynamic complementarity can also interact with the timing of borrowing constraints in subtle ways (Caucutt and Lochner (2017)). Carneiro and Heckman (2003) argue, the inability of children to borrow money to buy themselves out of family and neighborhood disadvantage is a kind of market failure that plays a critical role in their labor market outcomes. The importance of timing is not limited to family investments but more generally extends to cover the impact of prenatal and early childhood environments on long run outcomes (Almond and Currie (2011)). Heckman and Mosso (2014) summarize the evidence on the divergence in skills from early childhood and how these gaps play a decisive role in determining the life course social and economic outcomes of an individual. These reasons suggest that the standard IGE approach that proxies the permanent income with a simple average ignores important intergenerational mechanisms at work.[2]

In this chapter we take a different approach. We argue that the notion of permanent income as it was conceptualized by Friedman (1957) was not statistical but rather behavioral. In fact, broadly defined, the permanent component of income can be viewed as the effect of the factors that the individual perceives as the determinants of her life-time nonhuman and human wealth including personal attributes (abilities, personality, experience, occupation), family characteristics (e.g., marital status), location (e.g., schools, economic activity, social capital). Hence, there is no apriori reason to believe that a simple average of observed income is a sufficient statistic for permanent income, which in turn can be used to measure intergenerational mobility. Therefore, we treat the annual income data as discrete signals of a latent income process, which allows us to measure mobility by intergenerational income trajectories over the ages of exposure of children and young adults. Put differently, we shift the focus of the analysis from scalars to curves that map the trajectory of parent's outcomes to the trajectory of child's outcomes. In doing so, we employ a functional regression approach that treats the observations as "snapshots" of an underlying latent curve to uncover trends and accelerations in the intergenerational trajectories that may be revelatory of the importance of the timing of shocks over the entire lifetime of an individual.

Remarkably, the literature has focused on the linear IGE model and much of the discussion is limited to minimizing the problems of attenuation and life-cycle biases.[3]

[1] Mazumder (2005, 2015) argue that these biases can be addressed by constructing long time averages centered at age 40.

[2] These income shocks are generally endogenous due to parents' choices but this issue is beyond the scope of this paper.

[3] For example, Mazumder (2015) provides a systematic analysis of the two biases.

A notable exception is Carneiro, Italo, Salvanes, and Tominey (2018) who examine the role of the timing of parental income during their childhood years using administrative data from Norway. In particular, they estimate semi-parametric regressions of human capital outcomes on the discounted household income for the years when the child was between 0 and 17 years that proxies the permanent income, and two other measures of income at different stages of childhood that correspond to middle (ages 6-11) and late (ages 12-17). They find that when households income is shifted from middle childhood to either the early years or later childhood, the child can achieve optimal outcomes. Our analysis generalizes their work by considering labor market outcomes of the offspring using annual and bi-annual data that extend the age of exposure to include transition to adulthood (18-24) using a functional data approach. In the same spirit, Chausse, Chen, and Couch (2015) develop a multivariate functional regression method and provide a simple illustration of its usefulness in the context of intergenerational mobility. Yet, both of these analyses stop short from providing a full understanding of the intergenerational impact of the timing of shocks.

This paper builds on the aforementioned line of work by having the following three contributions. First, we examine the role of the timing of parental income during their childhood years in the US data based on PSID data. Second, we employ a functional data approach that allows the use of annual or bi-annual data and extend the time-span to include young adulthood. In contrast, the analysis by Carneiro, Italo, Salvanes, and Tominey (2018) is limited to three aggregate periods during childhood. Third, we investigate the heterogeneity in the intergenerational trajectories of income due to the socioeconomic status and structure of the family. There are several reasons why we should expect such heterogeneity in the intergenerational trajectories. One reason is the presence of intergenerational credit constraints by which we mean the inability of parents to borrow against the future income of offspring in order to invest in education (e.g., Loury (1981) and Galor and Zeira (1993), Han and Mulligan (2001)). Another possibility is the presence of intragenenational Aiyagari-type credit constraints that incorporate endogenous labor supply, human capital accumulation, and various psychic costs (Hai and Heckman (2017)). A third possibility is the presence of neighborhood effects that emphasize the importance of social factors in the intergenerational dynamics (Benabou (1996) and Durlauf (1996a,b)).

Our findings, based on intergenerational trajectories, suggest a richer and more nuanced characterization of the intergenerational mobility process than the standard empirical practice which is based on the IGE coefficient. In particular, not only we find that parental income in early and late childhood is important for their long run outcomes but also, parental income for young adulthood can be at least as productive as the ones in early and late childhood. More importantly, we provide ample of evidence that the shape of intergenerational trajectories may crucially depend on the socioeconomic background of the parents and family structure. Finally, we show that the shape of these trajectories is also sensitive to the timing of the shocks. Specifically, middle income and non-intact families exhibit higher sensitivity to shocks. Our results can be interpreted as suggestive evidence for the existence of complementarities in investments in human capital across periods in the spirit of Cunha, Heckman, and Schennach (2010).

The paper is organized as follows. Section 1.2 describes our data. Section 1.3 reviews the standard empirical approach that focuses on the linear IGE model. Section 1.4 presents the functional data approach and section 1.5 presents our results. Section 1.6 discusses future work and section 1.7 concludes and discusses future work.

1.2 Data

The data are drawn from the Panel Study of Income Dynamics (PSID). PSID is a longitudinal household survey starting in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the United States. We use the Survey Research Center, which is nationally representative. Adult children are linked to parents regardless of whether the parents are biological or adoptive. Specifically, we use as parent's income the income of the person that the child is living with at each age.

In our empirical analysis we investigate two age of exposures sample periods: a shorter sample period provides the annual incomes of parents when their children were 1 to 18 years old and a longer sample period extends the coverage to 24 years old but sampled biannually.[4] Henceforth, we will refer to these sample as short sample and long sample, respectively. In the short sample the children were born between 1968 to 1979 while in the long sample they were born between 1968 to 1981. The need for considering two samples is due to the unavailability of data for older children. The effective sample size shrinks from 580 in the case of short sample to 212 in the case of the long sample. The years 19-24 are expected to capture parental investments and timing of events related to college education and other transfers during the transition of adulthood of their offsprings.[5] In that case, the size increases to 887 in short sample and to 685 in the long sample. Another reason we sample income biannual is the fact that after 1997 the survey is conducted biannually in PSID. As our baseline, we consider the annual short sample and the biannual long sample. When we consider heterogeneity, we use biannual short and long sample.

[4]When income is sampled biannually, we take the average of the discounted income by the age of child.

[5]We trimmed observations both at the bottom and at the top by 3% to account for outliers.

Child's income as an adult is measured by labor income as a 3-year average centered at 35 years old. Labor income includes labor part of farm income and business income, wages, bonuses, overtime, commissions, professional practice, labor part of income from roomers and boarders or business income; the use of child's labor income as opposed to child's family income emphasizes our focus on human capital investment and direct labor market outcomes and avoids possible implications of assortative mating (e.g., Mazumder (2005), Mazumder (2015) and Landers and Heckman (2016)).

Parent's income when the child is at age a is measured by family income discounted with the age of the child defined as the taxable income of all earners in the family, from all sources, and transfer payments in order to capture all the resources available for parental investments on the child. The transfer payments include amount of aid to dependent children, aid to dependent children with unemployed fathers (ADC, AFDC) for the Head and Wife and for the entire family, income of Head and Wife for other Welfare, from Social Security, other retirement pay, pensions or annuities, from unemployment, or workmen's compensation, from alimony or child support, help from relatives, head's income from other sources, other transfer income of wife and transfer income from others in family. The ages of the parents vary from the ages of 16 to 70. For example, in the case of the short sample, when the child is at age 1 parents are on average 27 years old and when the child is at age 18 parents are on average 44 years old. All income measures are converted to 2011 dollars using the Consumer Price Index and adjusted for family size, by dividing with the square root of the family size. Finally, we compute the logarithm of both parents' income and child's average income.[6].

In measuring permanent income one has to deal with two challenges. Single-year measures of parental incomes are subject to transitory variations and measurement error, that may result to downward bias. This problem is known as attenuation bias. A typical method to address this problem is to average parents' income over several years. Solon (1992) argued that by using a 5-year average of income instead of a single year of income the bias shrinks substantially. Mazumder (2005) argued that even using a 5-year average may lead to a bias because the transitory variance in earnings is highly persistent. A second challenge is the lifecycle bias, which refers to the timing of measurement of income, that is, the income will not be representative of the individual's life time income if it is measured at either too young or too old ages. This bias can occur for at least two reasons (Mazumder (2015)). First, individuals with higher permanent income often have steeper income distributions that those with lower income and second, transitory fluctuations typically are much higher when the individuals are either too young or too old. Haider and Solon (2006) and Mazumder (2015) show this bias is minimized around 40.

[6]Negative or zero parental income values are set to one, while zero labor incomes are set to missing.

Our focus on the age of exposures shifts the focus from measurement issues to modeling issues but also imposes restrictions on how we address the above challenges. In particular, in this chapter we are interested in modeling the intergenerational trajectories of income using annual income observations for up to 24 years for the parents and 3 years for the children centered at the age of 35.[7] As we discuss below our method explicitly takes into account the fact that these observations are measured with error. In addition to the annual income we also use the annual stock of income defined as the annual cumulative income of an individual. This measure reflects the idea that if timing matters then parental investments can be better captured by the stock of human and asset wealth at any point of time. In fact the idea of using cumulative income and inputs is often used in the human capital literature (e.g., Bernal and Keane (2011)). There is also a statistical reason for using the stock of income variable. This variable is intrinsically continuous at any point of time and hence, it is expected to have more information than the annual income.

Beyond income we use offspring's and father's years of schooling and educational attainment. Educational attainment is measured by years of completed schooling for *High School Graduates* and *College Graduates*. For family structure we consider intact and non-intact families. An intact family is defined as a family for which both parents stayed together for the entire childhood of the offspring until the age of 18 or until the offspring left home to create her own household; whichever occurs earlier. A non-intact family is then defined as all the other types of families including blended and single-parent families.[8] Finally, as an additional offspring's outcome we use complete years of education. We include individuals that are at least 25 years old, in an age that typically educational attainment is measured.

Table 1.1-Panel A presents descriptive statistics for the baseline short and long sample when we consider permanent labor outcome as dependent variable. Figure 1.1 presents the income profiles for parent's income for the baseline samples. They show that parent's income increases until the child is around 20 years old and then stabilizes with a small decrease at age 24. Table 1.1-Panel B presents descriptive statistics for the baseline short and long sample when we consider father's years of schooling as an additional regressor. Table 1.1-Panel C presents descriptive statistics when the offspring's outcome is completed years of schooling. In that case, in the short sample we have 820 individuals who were born between 1968 to 1979 when we consider annual parental incomes. When we consider biannual samples, the size increases to 1907 in short sample and to 1392 in the long sample and the individuals were born between 1968 to 1988.

[7]We have chosen a three-year average in order to maximize our sample; a five-year average was also considered but the sample size would not allow us to make a comprehensive analysis.

[8]Unfortunately our small sample size does not allow us to distinguish among the various types of shocks that generate non-intact families such as death in the family, divorce, child out-of-wedlock, etc.

1.3 The standard approach

The standard empirical model focuses on intergenerational elasticity of income (IGE) given by the coefficient β in the following linear regression model,

$$y_{o,i} = \alpha + \beta y_{p,i} + \gamma' x_i + e_i, \tag{1.1}$$

where $y_{o,i}$ and $y_{p,i}$ denote the logarithms of the permanent incomes of the offspring and parent, respectively. These permanent incomes are proxied by simple averages, $y_{o,i} = \frac{1}{S}\sum_{s=1}^{S} y_{o,i,s}$ and $y_{p,i} = \frac{1}{T}\sum_{t=1}^{T} y_{p,i,t}$, where $y_{o,i,s} = log(Y_{i,t})$ and $y_{p,i,t} = log((1+r)^{-t}Y_{p,i,t})$, respectively.[9] x_i is a $k_x \times 1$ vector of other controls, typically involving age and age-squared that account for life cycle considerations when measuring permanent income. e_i is the regression error. As discussed in the introduction the magnitude of the IGE coefficient β determines the degree of intergenerational mobility. An IGE close to zero implies greater mobility while an IGE value close to one implies higher degree of immobility across generations.

One problem with the IGE model (1.1) is that it ignores the parental influence during the ages of exposure of children and young adults. A naive generalization of (1.1) that allows for such exposure effects is given by

$$y_{o,i} = \alpha + \sum_{t=1}^{T} \beta_t y_{p,i,t} + \gamma' x_i + e_i, \tag{1.2}$$

This model embodies effects during the ages of exposures when $\beta_1 = ... = \beta_T \neq 0$. This hypotheses can be tested using a joint Wald test.[10] Alternatively, under the normalization $\sum_{t=1}^{T} \beta_t = 1$ equation (1.2) can be rewritten as the sum of average income and future higher order differences

$$y_{o,i} = \alpha + \beta y_{p,i} + \beta \left(\sum_{j=1}^{T-1} \beta^*_{T-(j-1)} \Delta^{T-j} y_{p,i,T-(j-1)} \right) + \gamma' x_i + e_i \tag{1.3}$$

where $\sum_{j=1}^{T-1} \beta^*_{T-(j-1)} = 0$. This equation implies that the traditional linear IGE regression which only conditions on a proxy of permanent income will result in omitted variable bias when future higher order differences are correlated with the proxy of permanent income.[11]

[9] While the discounting is typically ignored in the literature, we use it for it is closer to the theoretical notion of permanent income. We assume $r = 0.03$.

[10] Note that (1.2) and the standard IGE non-nested are not nested due to the logarithmic transformation. For robustness purposes we also explore the model $y_{o,i} = \alpha + \beta log\left(\sum_{t=1}^{T} \beta_t (1+r)^{-t} Y_{i,t}\right) + \gamma' x_i + e_i$. Under $H_0 : \beta_1 = ... = \beta_T = 1/T$ we obtain the standard IGE.

[11] Note that this formulation as well as the DWH test applied to equation (1.4) can be viewed as a special case of Andreou, Ghysels, and Kourtellos (2010) who studied mixed frequency models when a

Table (1.2) provides evidence that the standard IGE model that emphasizes the permanent income as the driving force of long run outcomes and ignores the childrearing and developmental trajectories is not supported by the data. In particular, viewing the problem of testing for the standard IGE model as a problem of omitted variables we can view the annual income variables and their higher order differences as instrumental variables $z_i \in \{\{y_{p,i,j}\}_{j=1}^{T}, \{\Delta^{T-j} y_{p,i,T-(j-1)}\}_{j=1}^{T}\}$. We can then test for the null of no omitted variable bias (i.e., standard IGE model) by testing for $H_0 : \delta = 0$ using a standard Durbin-Wu-Hausman (DWH) test in the auxiliary regression

$$y_{o,i} = \gamma_0 + \gamma_1 y_{p,i} + \delta \hat{v}_i + u_i^*, \tag{1.4}$$

where $\hat{v}_i = y_{p,i} - \pi' z_i$. Table (1.2) provides p-values for the above LM test using both the short and long sample and two kinds of instruments. Columns 2 and 3 use as instrumental variables the annual parental income variables at age j one-at-a-time while Columns 4 and 5 use as instrumental variables the cumulative income up to age j, $\sum_{k=1}^{j} y_{p,i,j}$. We see that there is ample of evidence that the IGE model (1.1) is not supported by the data. In particular, the evidence against the IGE model is strongest for the case of stock and especially for the ages of middle and late childhood.[12]

Next, we propose a methodology that shifts the focus from IGE coefficients to intergenerational mobility curves that can capture the life course trajectories of an individual.

1.4 A functional approach

In this section we propose to study patterns of intergenerational mobility using a functional data approach that explicitly model trajectories. Traditional methods are limited in their capacity to capture the dynamics of functions. For example, if we want to study how the annual changes in family resources predict changes in offspring outcomes the linear IGE model (1.1) is unsuitable. The naive model (1.2) is also problematic because it does not model the trajectories as curves but rather as distinct parameter estimates rendering the interpretation of the results difficult. As a result this approach is subject to the incidental parameter problem since the parameters of the model increase with the number of years of exposure.

The idea of functional data analysis is to treat data as discrete measurements of an underlying continuous smooth stochastic L^2 process. Note that functional data are intrinsically infinite-dimensional and hence, subject to the curse of dimensionality. This

high-frequency variable is used to predict a low-frequency variable.

[12] While the results that include age effects are a bit weaker they do not alter the main conclusion that the linear IGE is not supported by the data. Those results are available upon request.

challenge is overcome by assuming smoothness such as the existence of continuous second derivatives, which is in fact a regularization assumption as it allows measurements at neighboring time points to be combined. Smoothing also allows handling missing values, sparse longitudinal data, and for measurement error by exploiting the repeated measurements for each individuals.

Specifically, the intergenerational income data are treated as n triplets $(y_{o,j}, y_{p,j}, t_j)$, $j = 1, ..., n$, where t_j lies in a bounded interval τ. $y_{o,ij}$ is assumed to be the ith signal of a smooth latent function $y_o(\cdot)$ so that $y_{oj} = y_o(\cdot) + u_{oj}$, where $y_o(\cdot)$ is random function and $u_j \sim i.i.d$ zero mean error. $y_{p,ij}$ can be defined likewise.[13] Ramsay and Silverman (1997) provide an excellent introduction on functional data analysis. Morris (2015) and Wang, Chiou, and Muller (2016) provide more recent surveys of estimation issues and applications of this approach.

In particular, using functional data analysis we can generalize the linear IGE model (1.1) to the intergenerational trajectories model

$$y_{o,i} = \alpha + \int_{\tau} \beta(t) y_{p,i}(t) dt + \gamma' x_i + e_i, \tag{1.5}$$

where $\beta(t)$ is the intergenerational elasticity function, which captures the intergenerational effect of the resources available to child at age t.

Furthermore, functional regression analysis allows for both functional offspring outcomes (dependent variable) and functional parental outcomes (explanatory variables) to vary over periods in highly nonlinear way allowing for measurement errors. Thus, unlike the naive model (1.2), the functional regression approach can capture how intra- and inter-offspring and parental outcomes coevolve. Assuming we can observe child's income over multiple periods we can also allow for a functional dependent variable using a function-on-function functional linear regression

$$y_{o,i}(s) = \alpha(s) + \int_{\tau} \beta(t,s) y_{p,i}(t) dt + \gamma' x_i + e_i(s) \tag{1.6}$$

The above intergenerational trajectories models can be generalized in a number of ways that can provide more insights about the patterns of intergenerational mobility by accounting for both static and dynamic heterogeneity. We discuss these extensions after we present our baseline results.

[13]More generally, the functions are not required to have the same set of argument values for all replications of each function, $i = 1, ..., n$, that is, (y_{ij}, t_{ij}), $j = 1, ..., n_i$ with $y_{ij} = y_i(\cdot) + u_{ij}$.

1.4.1 Estimation issues

Estimation of the functional regression model (1.5) is intrinsically an infinite dimensional problem. Typically, estimation is based on series estimation methods using splines or Fourier or principal components. Here we opt to use B-spline basis due to their computational speed and stability given the nature of our data. In particular, we apply smoothing both on the functional coefficients $\beta(t)$ as well as the regressor, in our baseline case, the parent's income $y_{p,i}(t)$. Smoothing $\beta(t)$ reduces collinearity in the regression, and makes the estimates more efficient and more interpretable while smoothing $y_{p,i}(t)$ reduces the measurement error of $y_{p,i}(t)$ and allows for sparse and irregular data.

More precisely, let $\{\phi_1(t), \phi_2(t), ...\}$ be a sequence of B-spline basis functions in L_2 space defined over a time interval $\tau = [t_0, t_T]$. We assume that $\beta(t)$ and $y_{p,i}(t)$ are approximated by the following B-spline basis functions

$$\beta(t) = \sum_{k=1}^{K_\beta} b_k \phi_k(t) = b' \phi_{K_\beta}(t) \tag{1.7}$$

and

$$y_{p,i}(t) = \sum_{l=1}^{L} c_{i,l} \phi_l(t) = c_i' \phi_L(t) \tag{1.8}$$

where $\phi_{K_\beta}(t) = (\phi_1(t), \ldots, \phi_{K_\beta}(t)(t))'$ and $\phi_L(t) = (\phi_1(t), \ldots, \phi_L(t)(t))'$ are $K_\beta \times 1$ and $L \times 1$ vectors of B-spline basis functions, respectively. b_i and c_i are the corresponding vectors of B-spline coefficients.

Given that B-splines bases are not necessarily orthonormal define $J_{kl} = \int_\tau \phi_k(t)\phi_l(t)' dt$. Then we can estimate (1.5) by least squares.

$$S_n(\alpha, b, \gamma) = \sum_{i}^{n} \left(y_{o,i} - \alpha - \sum_{k=1}^{K_\beta} \sum_{l=1}^{L} c_{i,l} J_{kl} b_k - \gamma' x_i \right)^2 . \tag{1.9}$$

Regularization is imposed by choosing the number of basis functions K_β and L which should depend on the sample size. We choose the number of basis functions using the generalized cross-validation (GCV) method, which is more reliable than cross-validation which tends to under-smooth.[14] Finally, we provide inference using bootstrap confi-

[14] We also explored an estimation method regularization with a roughness penalty (e.g., Cardot, Ferraty, and Sarda (2003)) based on the second derivative that aim at avoiding excessive local roughness with similar findings. To this end, the penalized LS criterion is given by $S_n^*(\alpha, b, \gamma, \lambda) = S_n(\alpha, b, \gamma) + \lambda \int_\tau (D^2\beta(t))^2 dt$ where D^2 is a linear differential operator and $\lambda > 0$ is a smoothing parameter that controls the trade-off between roughness and smoothness chosen by cross-validation. Our investigations did not yield substantially different results at least in the set of models we investigated so far.

dence intervals using 1000 replications.

1.5 Results

1.5.1 Income trajectories

We begin our analysis by presenting our baseline findings. Figures 1.2(a)-(b) show the estimates of intergenerational elasticity function $\hat{\beta}(t)$ of model (1.5) for both the short and long samples.[15] These estimates are plotted together with 90% wild bootstrap confidence intervals. The short sample covers three childhood periods of annual parent's income: early (1-6), middle (6-12), and late (12-18). The long sample extends the period to include the transition to young adulthood (18-24) with the difference that parent's income is sampled biannually.[16]

In general, our results show that the intergenerational elasticity function exhibits evidence of temporal heterogeneity (nonlinear pattern) for the ages of exposure following a subtle U-shape. This heterogeneity is strongest and with significant effects mainly for the long sample. Nevertheless, the prevailing pattern in Figures 1.2(a)-(b) is that parent's income is more productive in the early and late childhood. This finding is consistent with the finding of Carneiro, Italo, Salvanes, and Tominey (2018) who find that parent's income in the early and late childhood years is more important for the offspring's educational outcome than parent's income in the middle years. Our results show that the importance of the late childhood carries over into young adulthood.[17] While we find that the importance of early years is stronger in the long sample than the short sample, in both cases $\hat{\beta}(t)$ appears to be decreasing and reaches its minimum in the middle childhood. This evidence supports the findings of the literature on the human capital formation and the evolution of skills that emphasizes the importance of early human capital investments (e.g., Cunha and Heckman (2007) and Cunha, Heckman, and Schennach (2010)).

More insights about the life course dynamics can be obtained by considering the stock of income that can be used to generate human capital and wealth. Recall that an individual generates income via the labor market based on her level of human capital. Also note that at any given age throughout childhood and young adulthood, offspring's

[15] We present our results without controlling for the age of parents since in general, ages are not significant and controlling for them does not change our main results, substantively.

[16] For comparison purposes, Figure A1 of the Appendix shows the corresponding figures for the bi-annual short sample.

[17] Figure A2 of the Appendix investigates the impact of the added individuals in the long sample by plotting the trajectories until the age of 18 for both analysis 1 to 18 (upper panel) and 1 to 24 (upper panel) and keeping the same number of individuals.

human capital is a partial weighted sum of all current and past parent's of some function of human capital investments. If parent's annual income determines the human capital investments at any given age then we should expect that offspring's income is determined by the trajectory of the cumulative parent's income of parents

$$y_{o,i} = \alpha + \int_{\tau} \beta(t)\tilde{y}_{p,i}(t)dt + \gamma' x_i + e_i, \tag{1.10}$$

where $\tilde{y}_{p,i}(t) = \sum_{j=0}^{t} y_{p,i,j}$. $\beta(t)$ captures the intergenerational trajectory of exposure to the stock of the resources available to child at age t. If the life-time human capital and wealth of the parents is the only thing that matters then we would expect that only the stock of the last period to matter. We also compute the partial effect $\partial y_{o,i}/\partial \tilde{y}_{p,i}(t) = \int_t^T \beta(j)d(j)$ which captures the current as well as all future effects of the age of exposures effects.

Figures 1.2 (c)-(d) present the corresponding findings based on equation (1.10). We see substantial temporal variation hovering around zero with an expanding variance for both samples. Interestingly, $\tilde{y}_{p,i}$ exhibits a strong upward trend for both short and long samples in their corresponding latter periods. This upward trend starts after the age of 15 for the short sample and after the age of 21 for the long sample.

Overall, we find that the parental income in early and late childhood is important for the outcomes of children as adults, consistent with the findings of Carneiro, Italo, Salvanes, and Tominey (2018), the literature on human capital development and complementarities in the human capital production function Cunha and Heckman (2007) and Cunha, Heckman, and Schennach (2010)), and income uncertainty and partial insurance (e.g., Blundell, Pistaferri, and Preston (2008)). Moreover, we find that parental investments for young adulthood can be at least as productive as the ones in early and late childhood. These findings provide evidence that income shocks that affect parental human capital investments in children can in turn affect their long run labor market outcomes. One interpretation of our findings is that the presence of income uncertainty and partial insurance can give rise to environments where investments in children respond to parental income shocks. Alternatively, in the presence of dynamic complementarities parents may find it optimal to shift resources from middle to early and late childhood as well as young adulthood.

1.5.2 Heterogeneity

In this section we investigate the heterogeneity of the trajectories with respect to parental income, parental education, and family structure. This type of analysis will reveal whether the age of exposure for long run outcomes matters differentially for different types of family environments. As discussed in the introduction this heterogeneity may be manifestation of borrowing constraints or neighborhood effects.

Specifically, we consider threshold-type regressions with pre-specified regimes (sub-samples)

$$y_{o,i} = \sum_{j}^{q} \left(\alpha_j + \int_{\tau} \beta_j(t) y_{p,i}(t) dt + \gamma_j' x_i \right) I(z_i = j) + e_i, \tag{1.11}$$

where $I(z_i = j)$ is an indicator function that takes the value 1 if $z_i = j$ and otherwise is 0.[18] Equation (1.11) partitions the baseline model (1.5) into q-disjoint regimes. We mainly focus on the results for the long sample but we also include in the Appendix results based on the short sample.

1.5.2.1 Parent's permanent income

We start by describing our results in Figures 1.4, 1.5, and 1.6 from sub-sample functional regressions where the regimes are determined by parent's permanent income quartiles using the long sample. Generally, the results show that individuals born to parents from different socioeconomic background have different trajectories. In particular, we highlight the following findings. First, Figure 1.4(a) and 1.5(a) show that children born to disadvantaged parents experience substantial negative exposure effects up until age 3 and profiles with very steep positive gradients. This finding is in contrast with how the trajectories of other quartiles behave during early childhood which start at positive values and follow decreasing trajectory. The substantial negative exposure effects in middle childhood suggest that disadvantaged parents would choose to transfer resources to late childhood since those investments are more productive. What is more striking is their trajectory during young adulthood which also exhibits a decreasing pattern as opposed to the other quartiles. Figure 1.6 illustrates this sharp difference in the partial effects of the trajectory of the stock of income. One possible justification of these results is the impossibility of parents to borrow against their child's future earnings. In this case parental wealth is a binding constraint and thus children coming from constrained families will have lower early and late investments (Becker and Tomes (1986)).

Second, the trajectory of the most advantaged group in Figure 1.5(d) appears to be the reflection of Figure 1.5(a) over the age-axis. There are substantial positive effects in early, middle, and young adulthood periods. Interestingly, the uptick during young adulthood becomes stronger by parental income suggesting the importance of parental

[18] One problem with this analysis is that it assumes an apiori partition of the sample. In other words, the classification is not data driven. A natural generalization is to use classification methods such as functional regression trees methods or functional threshold regression.

transfers for university education, which in turn will determine the labor market outcomes of the child. Figures 1.5(d) and Figure 1.3(c) clearly documents the importance of the stock of income during young adulthood.

Third, while the trajectories of the third quartile in Figure 1.4(c) exhibit higher uncertainty, it is worth noting that on average the third quartile exhibits higher persistence. This finding is consistent with Durlauf, Kourtellos, and Tan (2017) who find evidence of more persistence in earnings outcomes for members of middle income families.

Finally, we investigate the importance of the timing of family disadvantage by measuring the quartiles of permanent income using information only from early or late childhood. The relative impact of the timing of disadvantage can be attributed to the relative impact borrowing constraints or family shocks that change the family structure occurred in the early childhood rather than late childhood or young adulthood. Figures 1.7 and 1.8 show the intergenerational income and stock trajectories between early and late for the four quartiles, respectively. Figure 1.9 superimposes the partial effects for the four quartiles.[19] In general, as expected we find that the trajectories in the bottom of the income distribution are the least sensitive to the timing of the shock. Interestingly, the trajectories of the third quartile exhibit the most sensitivity. For instance, Figure 1.9(c) shows that trajectories of the partial effects of income stocks are much stronger for the ones whose parents experience late shock.

In sum, the results show that the socioeconomic background of the parents affect the intergenerational trajectories of the offspring. Children coming from the most disadvantaged background exhibit very different trajectories than other groups. There are considerable negative effects during middle childhood that indicates that parents choose to move investments from middle to late childhood since those investments are more productive. A possible explanation for these findings, is the existence of binding credit constraints. For children of the most advantaged parents, there are notable positive effects in early, middle, and early adulthood periods; especially in the early adulthood. Furthermore, consistent with the findings of Durlauf, Kourtellos, and Tan (2017), we find evidence of of relatively higher immobility in the middle of the income distribution rather than for the relatively disadvantaged and relatively affluent. More importantly, however, the degree of immobility is not only determined by socioeconomic class but also by the timing of parental income shocks.

[19] Figures A7 and A8 of the Appendix provide the corresponding figures for the short sample.

1.5.2.2 Father's education

One possibility is that higher levels of parental permanent income are associated with higher levels of parental education. Hence, we investigate heterogeneity with respect to the education level of father.

Figures 1.10, 1.11, and 1.12 present our results on intergenerational trajectories when we split the sample based on whether fathers graduated from high school or from college. Overall, our results are confirmatory of the heterogeneity by parent's income but weaker especially for the annual income trajectories.

In particular, the trajectories of non-high school graduates (Figures 1.10(a) and 1.11(a)) and non-college graduates (Figures 1.10(c) and 1.11(c)) appear to be similar to the trajectories of individuals whose parents had the lowest income. Likewise the patterns exhibited by the relative more advantageous groups in terms of income are similar to the trajectories of children whose father had at least high-school or college education.

Furthermore, Figure 1.12(a) shows that the trajectories of the partial effects of the stock income are always higher for the offsprings with fathers with at least high-school education relative to those with fathers who did not graduate from high-school. What is striking is the strong uptick for the offsprings with fathers with at least high-school education. This evidence suggests that father's education plays a key role in both early and late investments and thereby their long-run outcomes; see for example Keane and Woplin (2001) and Becker, Kominers, Murphy, and Spenkuch (2018).

Interestingly, comparing the same trajectories for offsprings with fathers with at least a college degree against those with fathers with non-college degree (Figure 1.12(b)), we find that the former sub-sample exhibits weaker effects possibly due to substitution effects.

In sum, we find evidence that father's education is important for the outcomes of the children as adults, through early and late childhood investments.

1.5.2.3 Family structure

Another source of heterogeneity in the intergenerational trajectories is due to family structure which influences the parent-child interactions (e.g., Moon (2014), Heckman and Mosso (2014)). Figures 1.13, 1.14, 1.15 show the income and stock trajectories for non-intact and intact family structures and the partial effects of the stock of income for the short sample.[20] For the non-intact sample we also distinguish between the individuals who experienced a family shock in their early childhood and those with a family shock in their late childhood.

The trajectories for the non-intact sub-samples exhibit interesting patterns. Focusing on Figure 1.13(b) that does not distinguish between the timing of the shock, we see a sharp decrease in the trajectory of income, becoming negative around the age of 3, and reaching a minimum at the age of 6. Then it becomes positive after age 10 and stabilizes in the late childhood. This pattern seems to be driven by individuals who experienced early shock as it is implied by the Figures 1.13(c)-(d). It is also worth pointing out that on average the partial effects of the stock of income for the children who experienced late shock are much higher than those who experienced early shocks. This suggests that those early family shocks tend to make the family investments less productive. Finally, we note that the trajectories of childhood exposures for the intact families exhibit a similar behavior as the bi-annual baseline (full) sample in Figure (A1).

Overall, the results indicates that early family shocks seem to affect more the parental investments, making them less productive.

1.5.3 Additional results

In this section we provide four additional investigations for robustness purposes. First, we redo the main analysis using child's schooling attainment as an outcome variable instead of child's income in Figures 1.16-1.19. Figure 1.16 shows the results for the intergenerational trajectories of income, stock of income, and partial effects of the stock of income. Not surprisingly, we find that the results using schooling attainment are generally similar to our main analysis that relies on child's income. For example, Figure 1.16(a) shows that the estimated curve exhibits the same subtle U-shape pattern as in the case of child's income in Figures 1.2(b). One notable difference, is the shape of partial effects of the stock of income for offsprings whose parent's permanent income lie in the first quartile in Figure 1.19. While in the case of income the trajectory of partial effects is decreasing from late childhood and turning even negative in adulthood, the corresponding trajectory in the case of schooling attainment is less decreasing for the same period, remains positive, and becomes increasing in the late young adulthood. This difference may be attributable to unobservables related to labor market shocks that affect the most disadvantaged offsprings.

[20] We opt to use short sample for family structure since the size of the non-intact subsample for the long sample is much smaller.

Second, in Figure 1.20 we investigate how our baseline results change if we analyze daughters and sons, separately. Interestingly, we find that the trajectory of parent's income for daughters appears to have an inverted U shape in the early childhood while for sons it appears to have a decreasing shape. Furthermore, we find that parent's income also appears to be important in the late childhood in both samples with the difference that in the long sample it follows an inverted-U pattern before it increases again during the young adulthood. This effect appears to be stronger for the daughters than the sons which may reflect the relative dependence of daughters on parents during young adulthood. This suggests that the time variation in the resources available to daughters has substantially more impact on their labor market outcomes than the corresponding effects for the sons.

Third, we take seriously a criticism that our analysis is likely to be highly correlated. Figure 1.21 investigates a model that considers the trajectory of income growth experiences $g_{p,i}(t) = y_{p,i,t} - y_{p,i,t-1}$ for $t = 2, ..., T$ conditional on initial income $y_{p,i,1}$. Consistent, with the findings based on parent's annual income and stock of income our results show that the trajectory of growth rates plays an important role for child's income for all ages of exposure and for both samples with the largest effects occurring in the early and late periods. Furthermore, we find that the coefficients of initial income are 0.40 and 0.42 for short and long samples, respectively. In both samples the coefficients of initial income is statistical significant at 1%, highlighting the importance of initial resources when a child is born.

Finally, we consider how the intergenerational trajectories respond to the inclusion of factors that may proxy the permanent income of the parents beyond annual income. Figure 1.22 of present results of equations (1.5) and (1.10) that condition on father's education. In general, we find our results remain qualitatively unaltered.

In summary, we find that our main conclusions are not altered when we use child's schooling attainment as an outcome variable, examine separately sons and daughters, consider different transformations of the parent's income such as growth rates, and condition on additional control variables such as father's education.

1.6 Future work

In terms of future work, an important and natural extension to our work is to investigate the role of dynamic complementarity, which means that human capital investments at different ages exhibit synergies and bolster each other (Cunha and Heckman (2007) and Cunha, Heckman, and Schennach (2010)). The idea of dynamic complementarity combines the ideas of self-productivity of skills and complementarity. An increase in current investments leads to an increase in next period's skills but because these skills exhibit self-productivity, in the sense that skills produced at one stage beget skills at later stages, current investments and future investments are always complements as long as future investments and future skills are contemporaneously complements. In fact, given that the marginal product of increasing investments is increasing in skills in any future period and the fact that future investments are complements with future skills, the degree of complementarity between current and future investments will be stronger, the stronger the future contemporaneous complementarity.

The ideas of dynamic complementarity are particularly important for policy making. If there exist such dynamic complementarities then policies that promote early human capital investments can have two implications. First, such policies make later policies that promote the formation of human capital more productive, and second, if early investments are not accompanied by later investments, those early policies are not effective. Cunha and Heckman (2007) summarize the empirical evidence on life cycle skill formation and present a model that accounts for a multistage technology of skill formation that features self productivity and dynamic complementarity.

There are two possible approaches that allows for modelling interaction between income at different ages. First, following Yao and Muller (2010), we could estimate model

$$y_{o,i} = \alpha + \int_{\tau} \beta_1(t) y_{p,i}(t) dt + \int_{\tau}\int_{\tau} \beta_2(t,s) y_{p,i}(t) y_{p,i}(s) dt ds + \gamma' x_i + e_i \tag{1.12}$$

An alternative semiparametric way would be to consider a varying coefficient functional linear regression model

$$y_{o,i} = \alpha(z_i) + \int_{\tau} \beta(z_i, t) y_{p,i}(t) dt + \gamma' x_i + e_i \tag{1.13}$$

where $z_i = (\bar{x}_{iearly}, \bar{x}_{imiddle}, \bar{x}_{ilate}, \bar{x}_{iadulthood})$ is the vector of average of parental income during early, middle, late childhood and early adulthood, respectively.

The above analysis will enable us to have some more insights of the policy implications of our empirical results. We have found that early and late investments are more productive than investments during middle childhood when the compulsory education starts (which starts between five and eight and ends somewhere between ages sixteen and eighteen, depending on the state). Our findings, along with the results of early interventions for children from disadvantaged families (e.g. Abecadarian Project, Perry Preschool experiment, Chicago child-parent programm) suggest that early investments are crucial for the human capital development of the offsprings. Ability gaps between individuals coming from different socioeconomic groups open up at early stages of life and remediation efforts in schooling years are not effective, since they do not eliminate these ability gaps (Hansen, Heckman, and Mullen (2004)).

Moreover, recent literature in the economics of human capital development establishes the importance of multiple skills distinguishing between cognitive and non cognitive skills, showing that earlier stages are crucial for the development of cognitive skills while later stages are crucial for the development of non-cognitive skills (Cunha, Heckman, and Schennach (2010)). Another important fact we want to investigate is that for the skill formation multiple forms for investments including parental time investments should be taken into account (Bernal (2008), Bernal and Keane (2010), Del Boca, Flinn, and Wiswall (2014)). Time investments might be substitutes or complements for goods investments, while spending time with children allows parents assess the abilities of their children and make more targeted investments. We are going to employ the Child Development Supplement (CDS) and Transition to Adulthood (TA) of the PSID data that include additional information on children and their parents including time use (diary) data, health, skills assessments, parenting styles, learning environment in the home, and socio-emotional characteristics of children and their parents. This rich database will allow us to consider models with functional outcome variables as well as generally provide a better understanding of the underlying mechanisms of intergenerational dynamics (e.g., Del Boca, Flinn, and Wiswall (2016), Caetano, Kinsler, and Teng (2017)).

Another avenue of future work is to consider the endogeneity of income shocks by modeling the timing of parental investments as in Cunha, Heckman, and Schennach (2010). Alternatively, the income variable can be decomposed into permanent and transitory components as in Carneiro, Salvanes, and Tominey (2016) and Abbott and Gallipoli (2019). Furthermore, a methodological extension of our work is to consider data functional mixture models (Yao, Fu, and Lee (2011)) and varying coefficient functional models (Wu, Fan, and Muller (2010) and Zhang and Wang (2015)) that can provide further insights for the presence of borrowing constraints and neighborhood influences. We are going to investigate the robustness of our results to other types of functional forms, e.g., orthonormal polynomials, as they could provide better properties in terms of efficiency. Moreover, we could consider other consistent information criteria beyond the GCV to choose the optimal order of the basis functions.

1.7 Conclusion

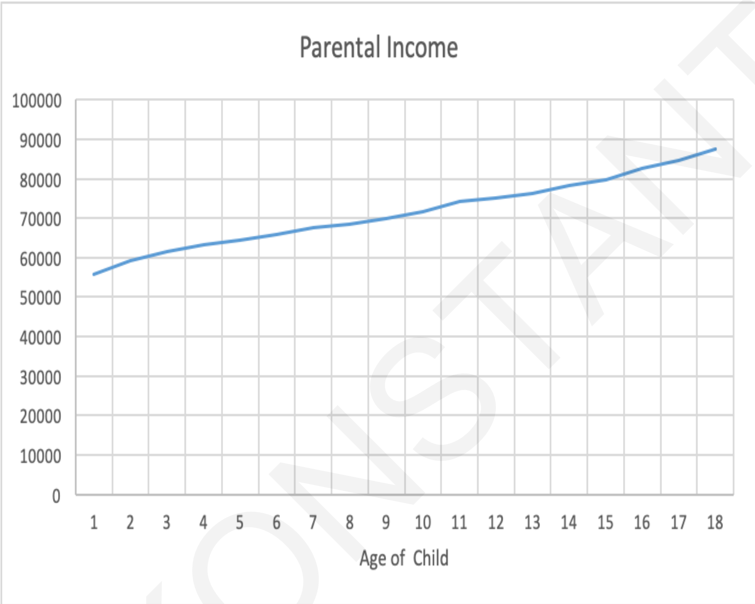
In this chapter we propose a novel way to measure intergenerational mobility of economic status. We argue that functional regressions provide a flexible and parsimonious way to capture the intergenerational effects of higher frequency influences during the age of exposure during childhood and young adulthood in ways not captured by the current empirical practice. We find that parental investments are generally more productive in the early and late childhood or young adulthood, suggesting that income shocks play a major role in parental human capital investments in children and in their long run outcomes. More importantly, we find that the timing of the shocks related to socioeconomic status and family structure can have a key role in the upward mobility of individuals, especially for disadvantaged children.

1.8 Figures

Figure 1.1: **Intergenerational Trajectories of Income**

This figure presents the average annual parental income per age of individual. Figure 1.1(a) presents the average annual parental income for the short baseline sample, while Figure 1.1(b) presents the average annual parental income for the long baseline sample.

(a) Short Sample



(b) Long Sample

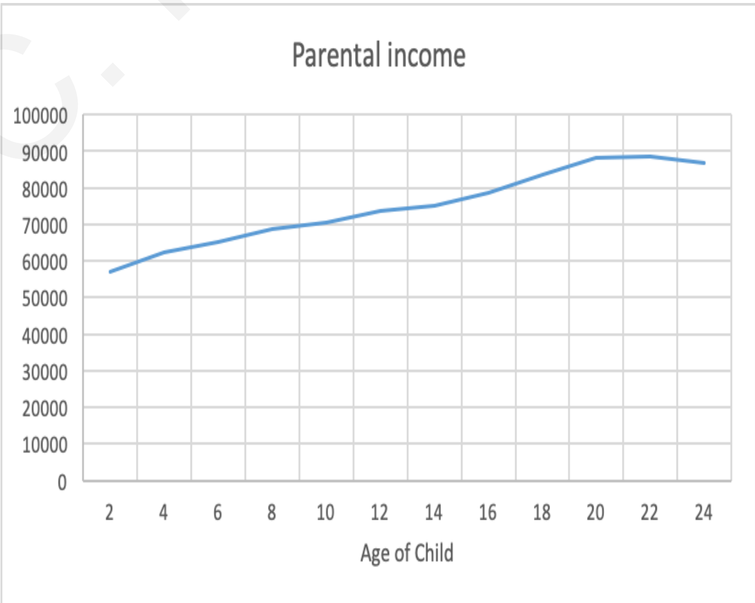
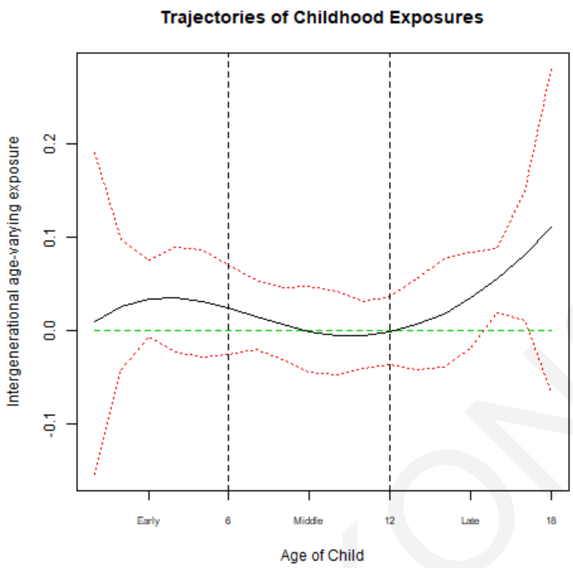


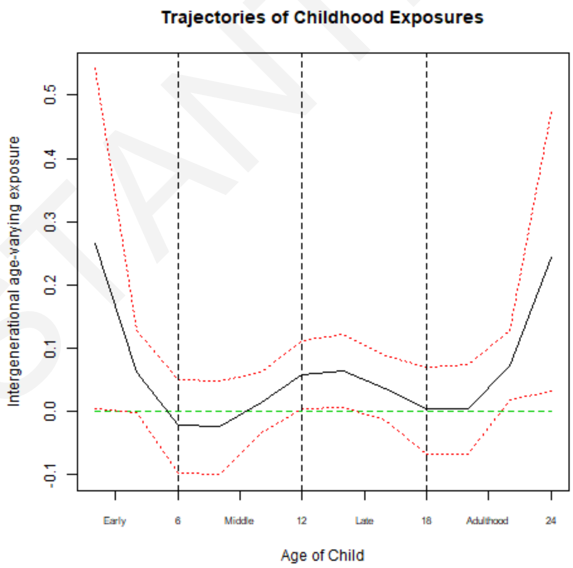
Figure 1.2: **Intergenerational Trajectories of Income**

This figure presents the baseline results from model (1.5). Figure 1.2(a)-(b) present estimates of the intergenerational elasticity of income $\hat{\beta}(t)$ for the short and long samples, respectively. Figures 1.2(c) and 1.2(d) present the corresponding functions for the stock of income from equation (1.10). The red dotted lines represent 90% bootstrap confidence bands.

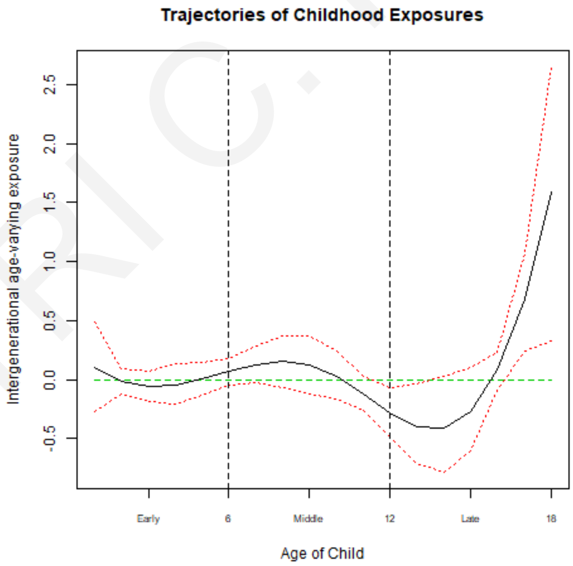
(a) Income - Short Sample



(b) Income - Long Sample



(c) Stock of Income - Short Sample



(d) Stock of Income - Long Sample

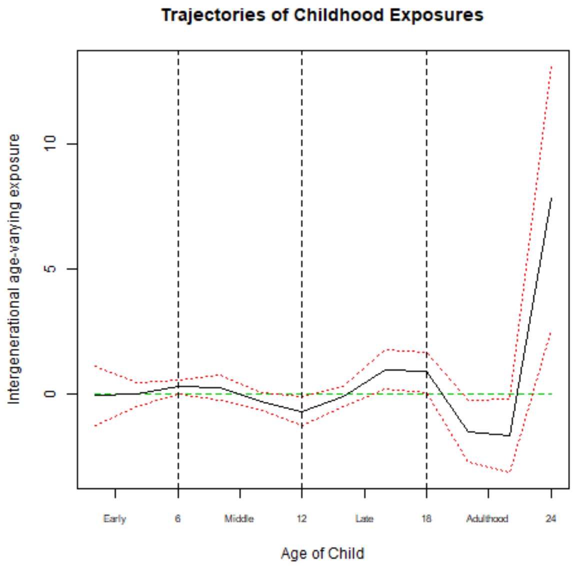


Figure 1.3: **Partial Effects of the Stock of Income**

This figure presents the trajectories of partial effects of stock of income in equation (1.10) for the short and long samples.

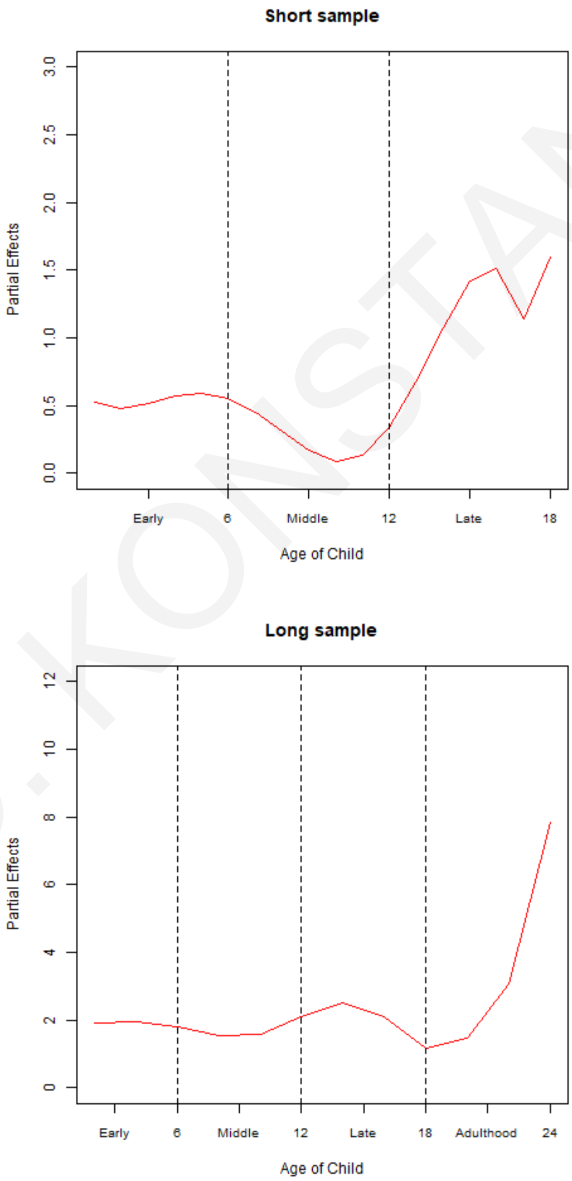
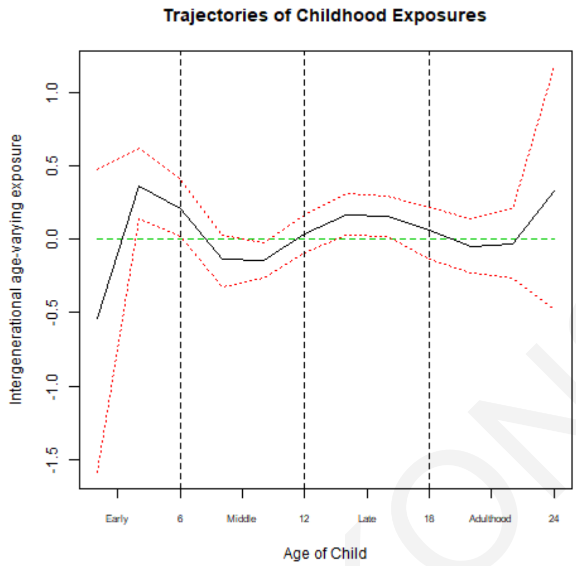


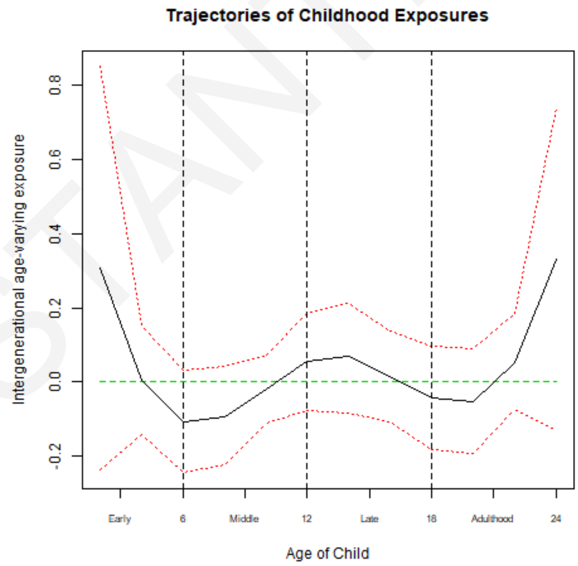
Figure 1.4: **Intergenerational Trajectories of Income by Parental Income Quartiles**

This figure presents the baseline results from model (1.5) for the long sample. Figures 1.4(a), (b), (c), (d) present the estimates of intergenerational elasticity function $\hat{\beta}(t)$ based on the long sample for the first, second, third and fourth parent's permanent income quartile respectively. The red dotted lines represent the bootstrap confidence bands.

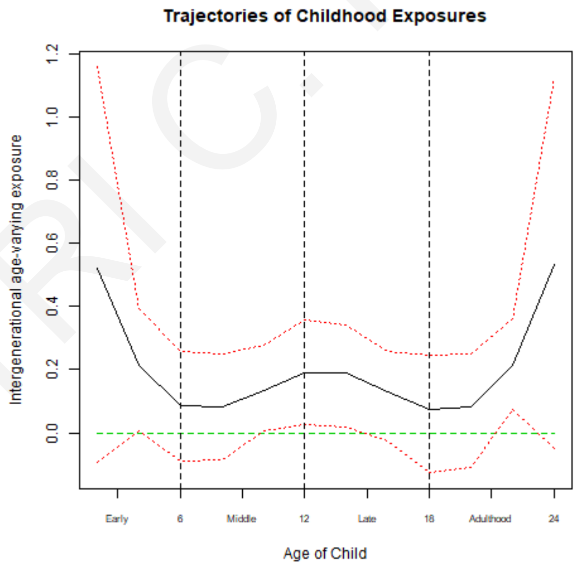
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

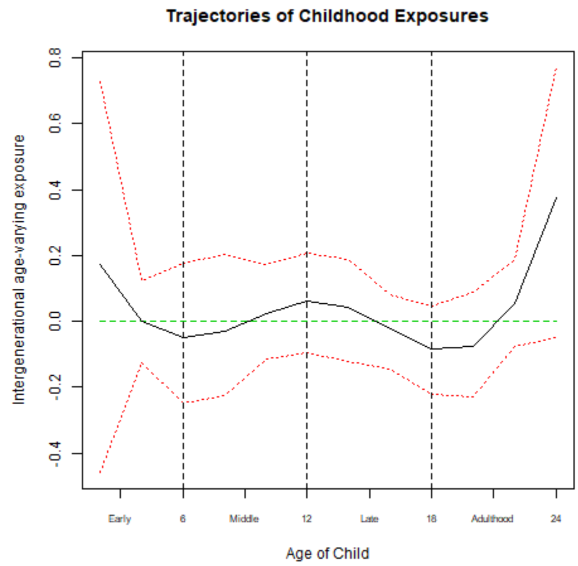
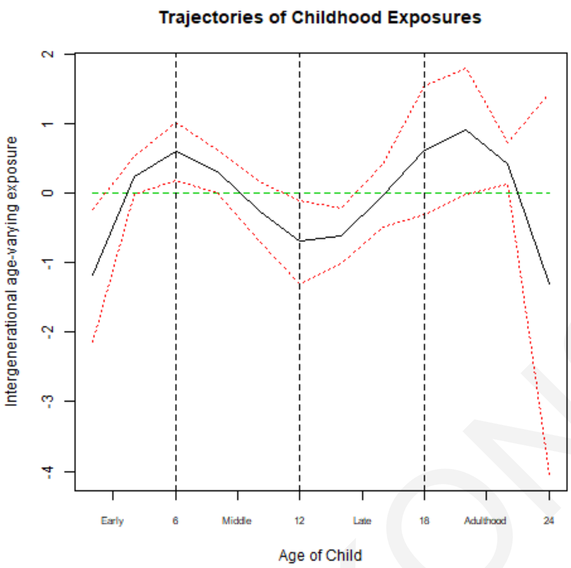


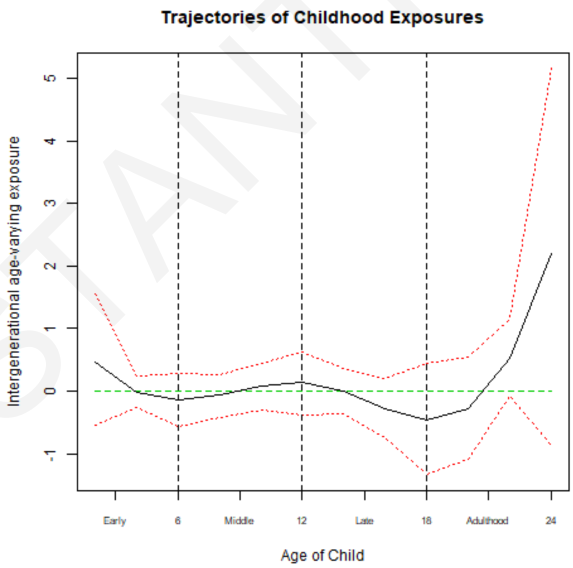
Figure 1.5: **Intergenerational Trajectories of the Stock of Income by Parental Income Quartiles**

This figure presents the baseline results from model (1.10), based on the long sample. Figures 1.5(a), (b), (c), (d) present the estimates for the first parent's permanent income quartile, second, third and fourth quartile respectively. The red dotted lines represent the bootstrap confidence bands.

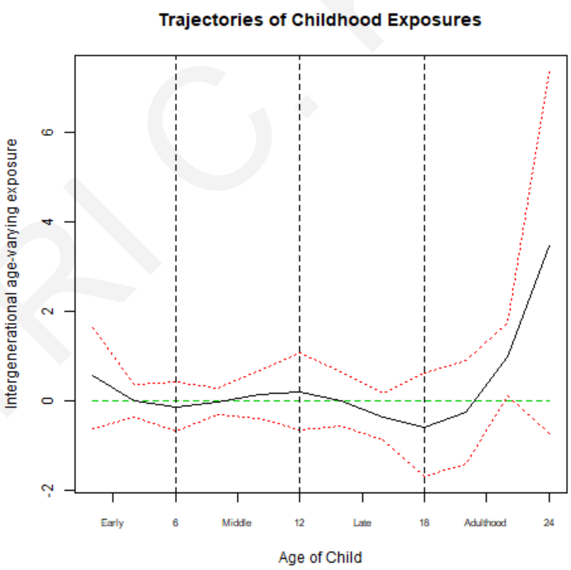
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

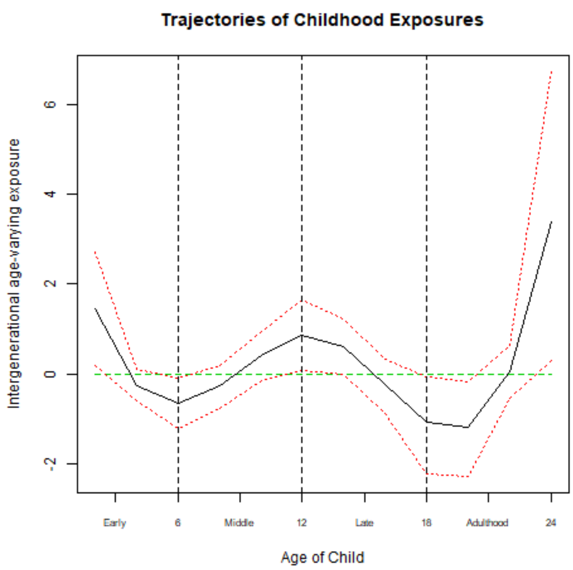


Figure 1.6: **Partial Effects of the Stock of Income based on Parental Income Quartiles**

This figure presents the trajectory partial effects of equation (1.10). The red line corresponds to first parental permanent income quartile, the green line to the second quartile, the blue line to third quartile and the cyan line to fourth quartile.

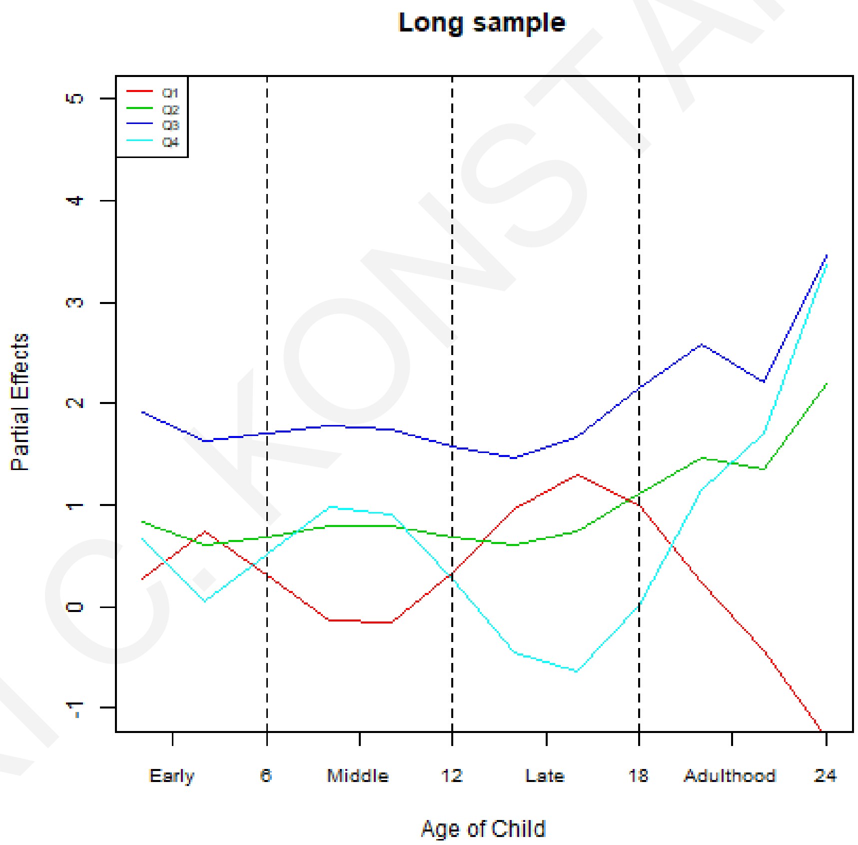
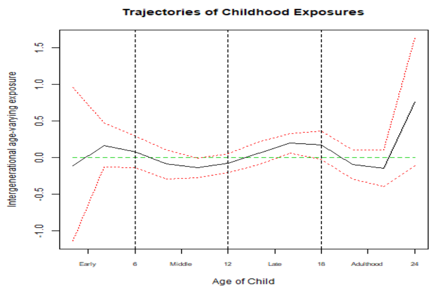


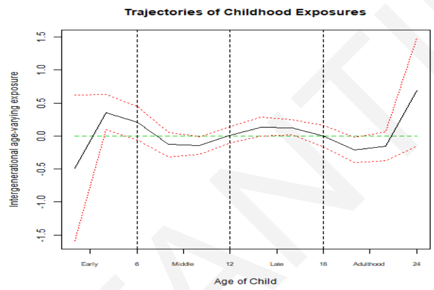
Figure 1.7: **Timing of Income Shocks - Trajectories of Income**

This figure compares the intergenerational trajectories of marginal effects for stocks of income for the long sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

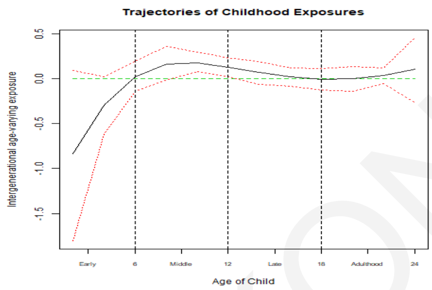
(a) First Quartile - Early



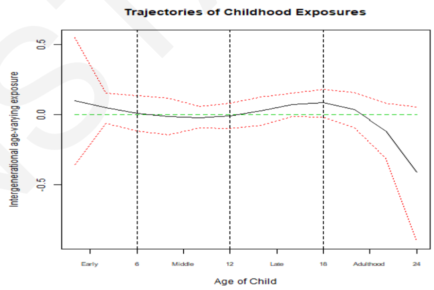
(b) First Quartile - Late



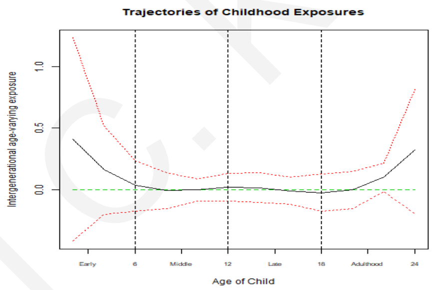
(c) Second Quartile - Early



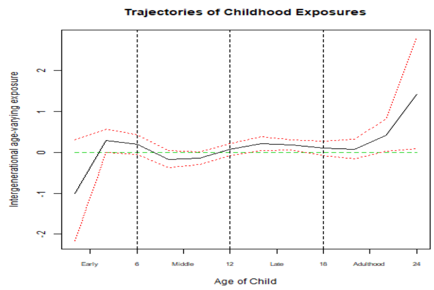
(d) Second Quartile - Late



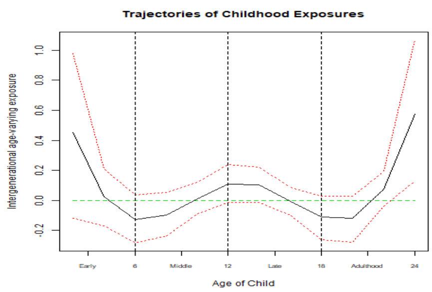
(e) Third Quartile - Early



(f) Third Quartile - Late



(g) Fourth Quartile - Early



(h) Fourth Quartile - Late

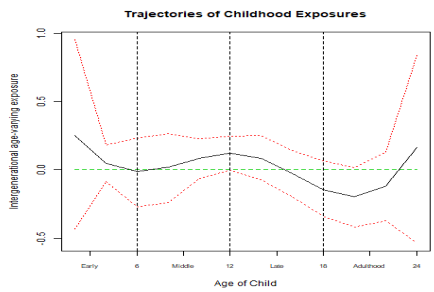
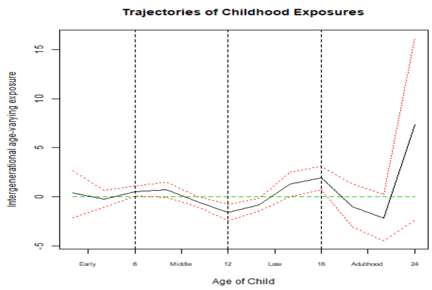


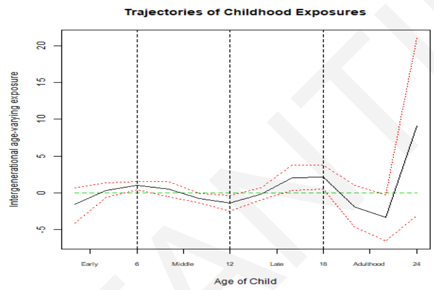
Figure 1.8: **Timing of Income Shocks - Trajectories of Stock of Income**

This figure compares the intergenerational trajectories of marginal effects for stocks of income for the long sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

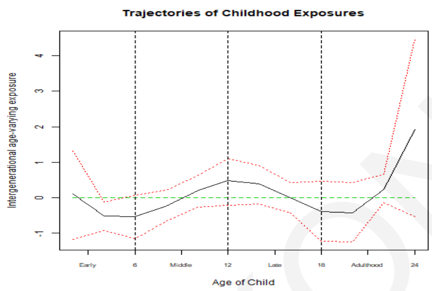
(a) First Quartile - Early



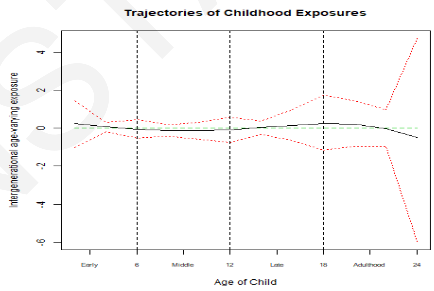
(b) First Quartile - Late



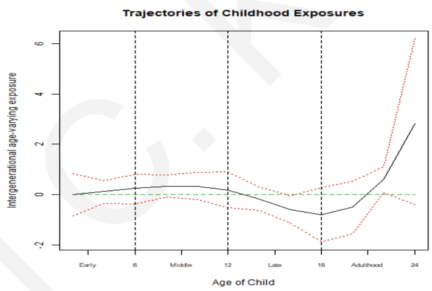
(c) Second Quartile - Early



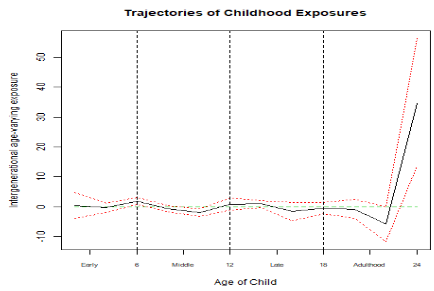
(d) Second Quartile - Late



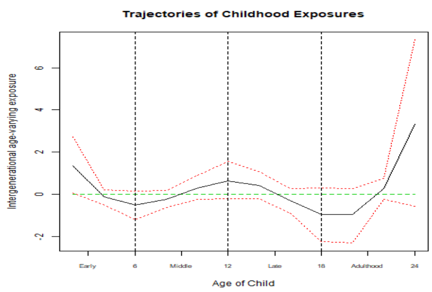
(e) Third Quartile - Early



(f) Third Quartile - Late



(g) Fourth Quartile - Early



(h) Fourth Quartile - Late

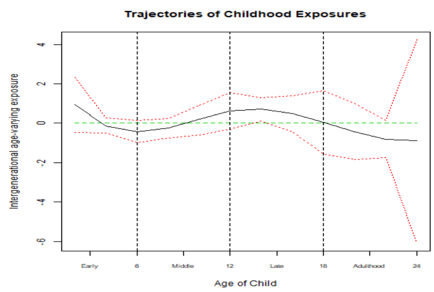
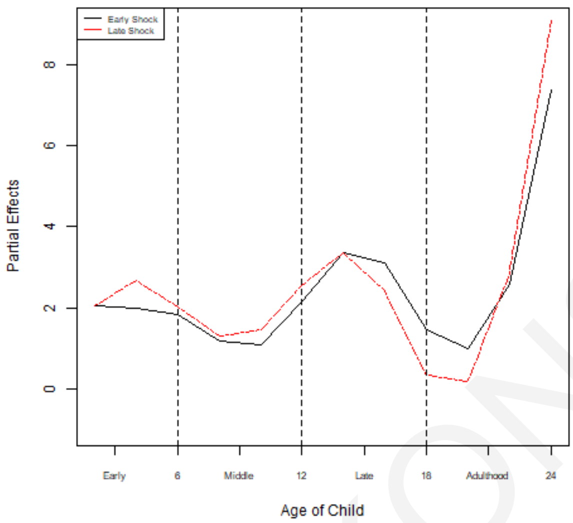


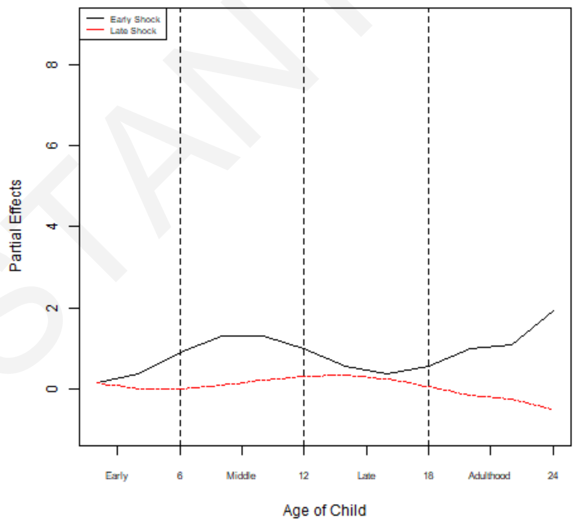
Figure 1.9: **Timing of Income Shocks - Trajectories of Partial Effects of Income**

This figure compares the intergenerational trajectories of marginal effects for stocks of income for the long sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

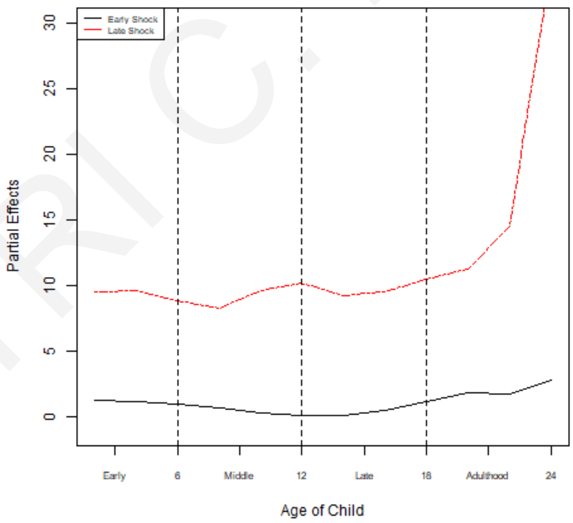
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(c) Fourth Quartile

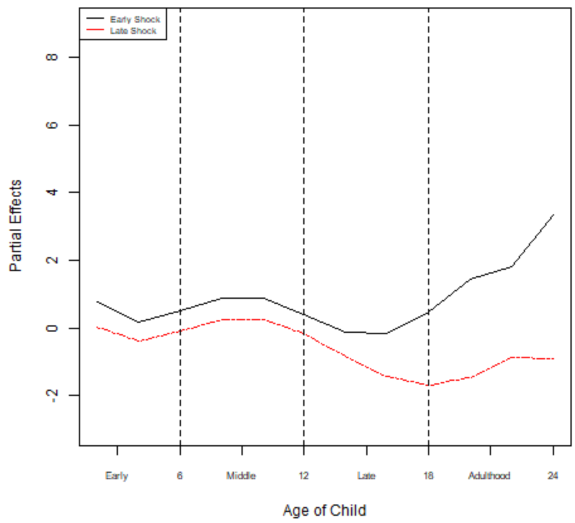
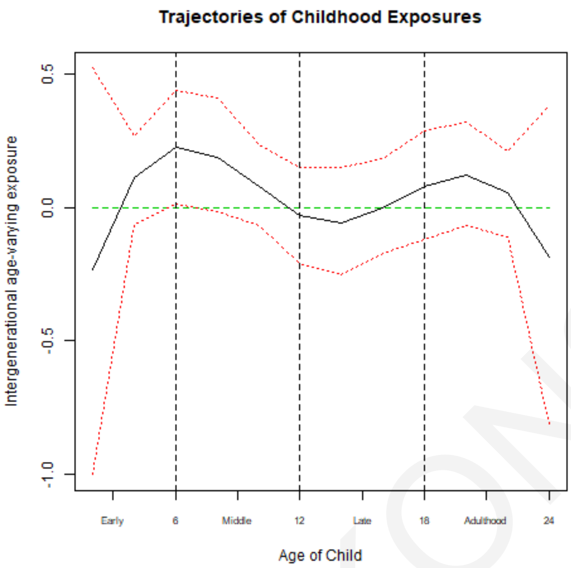


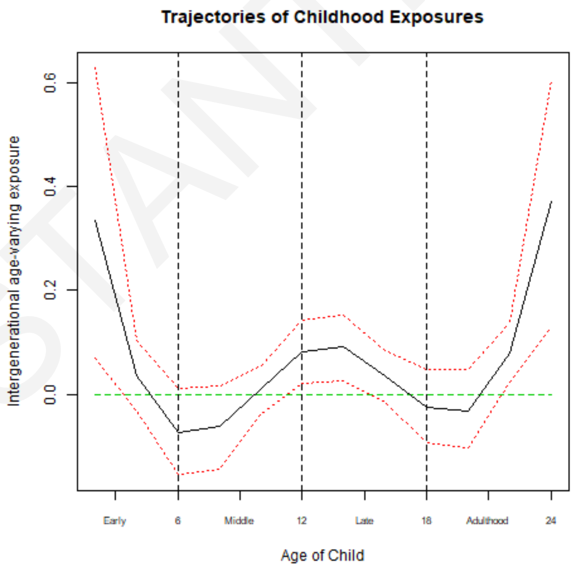
Figure 1.10: **Intergenerational Trajectories of Income by Father's Education**

This figure presents the baseline results from model (1.5) based on the long sample, for subsamples based on father's education. Figures 1.10(a)-(b) present the estimates of intergenerational elasticity function $\hat{\beta}(t)$ for individual's with non-high school graduates fathers and for individual's with high school graduates fathers, respectively, and Figures 1.10(c)-(d) for individual's with non-college graduates fathers and for individual's with college graduates fathers.

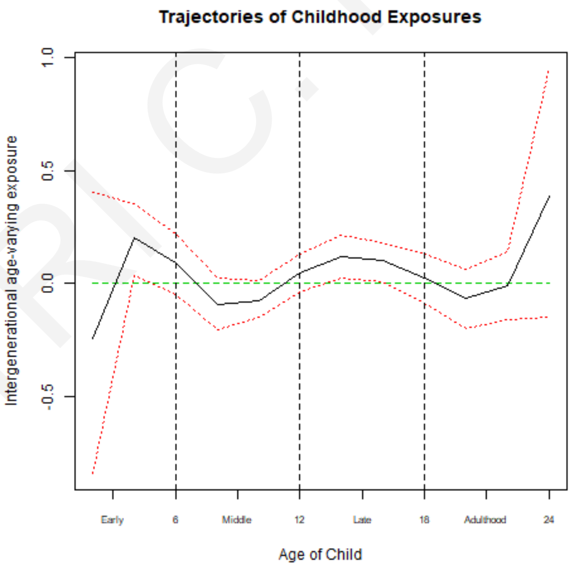
(a) Non-high School Graduates



(b) High School Graduates



(c) Non-college Graduates



(d) College Graduates

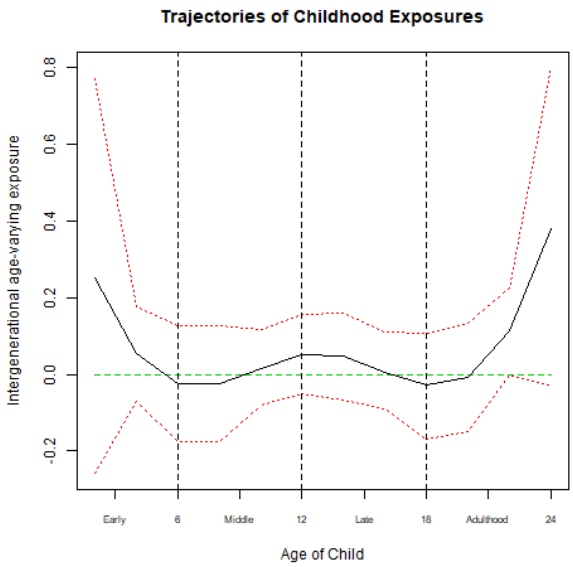
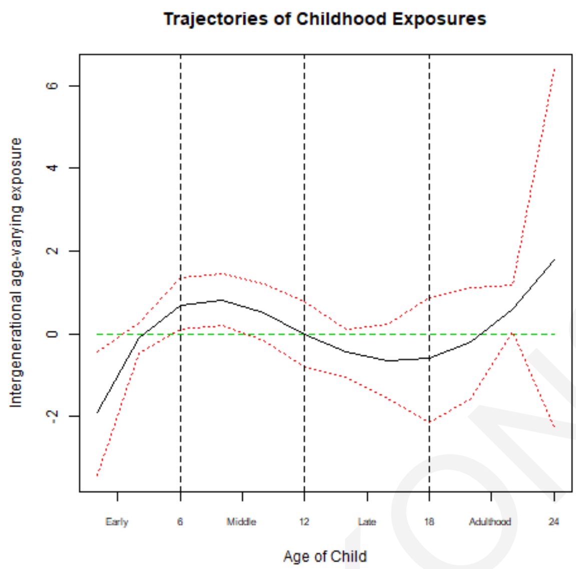


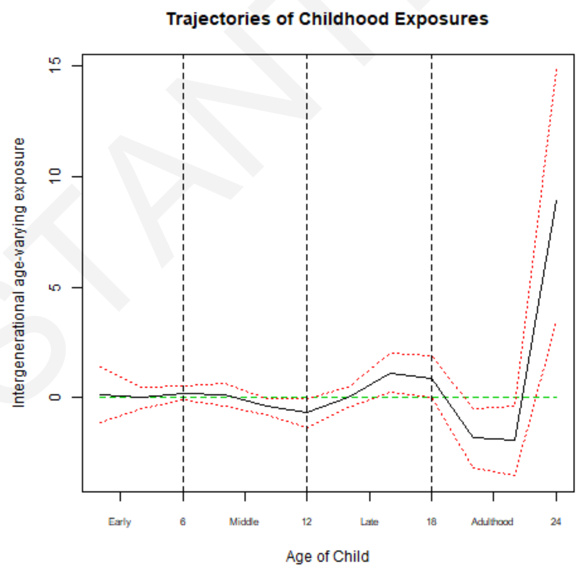
Figure 1.11: **Intergenerational Trajectories of the Stock of Income by Father's Education**

This figure presents the baseline results from model (1.10) based on the long sample, for subsamples based on father's education. Figure 1.11(a)-(b) present the estimates for individual's with non-high school graduates fathers and with for individual's high school graduates fathers, and 1.11(c)-(d) for individual's with non-college graduates fathers and for individual's with college graduates fathers.

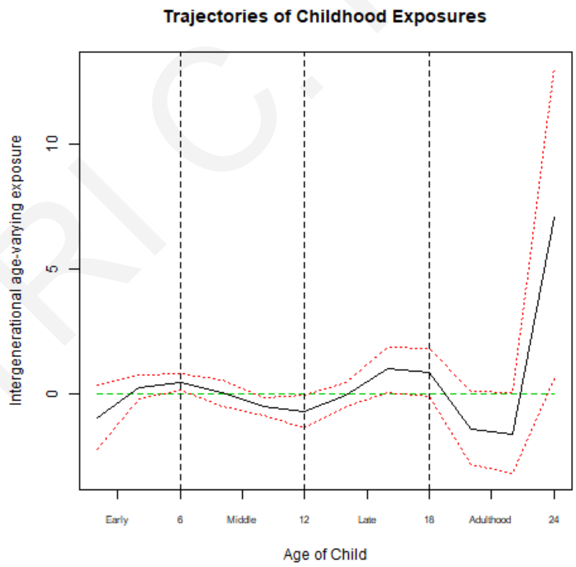
(a) Non-high School Graduates



(b) High School Graduates



(c) Non-college Graduates



(d) College Graduates

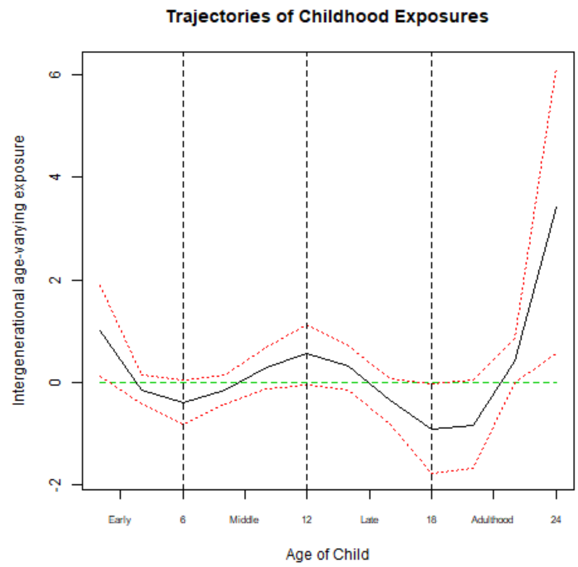
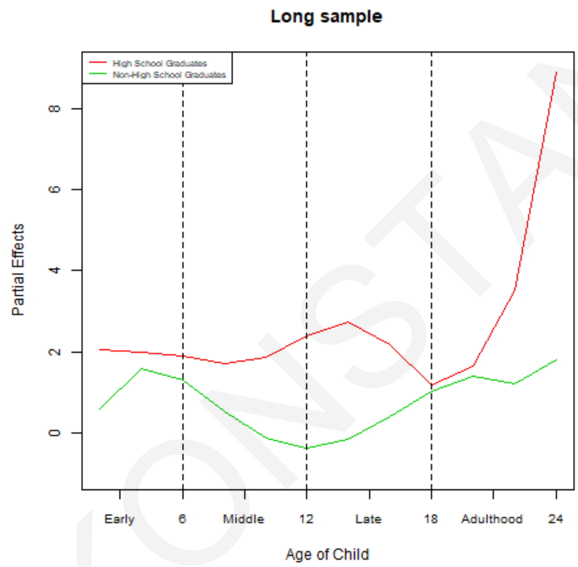


Figure 1.12: **Partial Effects of the Stock of Income by Father's Education**

This figure presents the trajectory partial effects of equation (1.10) based on the long sample, for subsamples based on father's education. Figure 1.12(a) presents the estimates of partial effects for individual's with non-high school graduates fathers with green line and for individual's with high school graduates fathers with red line. Figure 1.12(b) presents the estimates of partial effects for individual's with non-college graduates fathers with green line and for individual's with college graduates fathers with red line.

(a) Based on High School Graduation



(b) Based on College Graduation

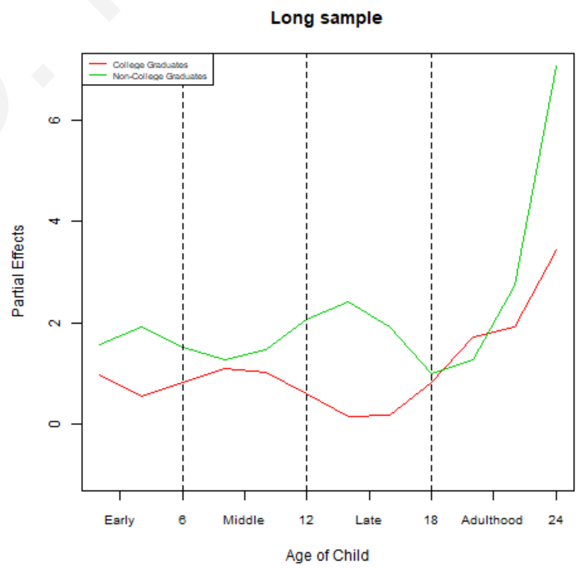
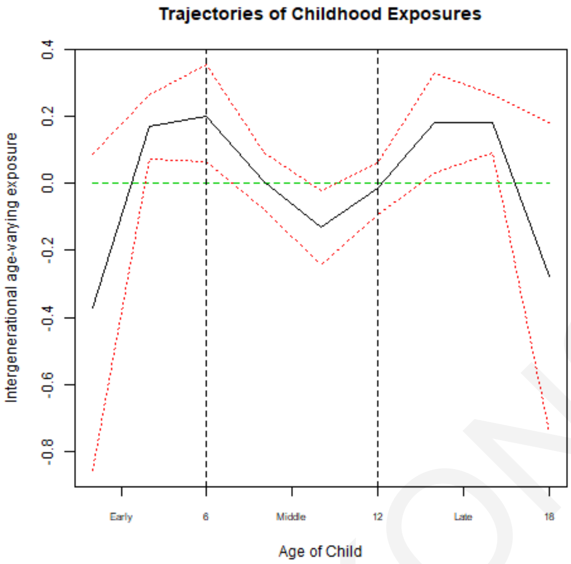


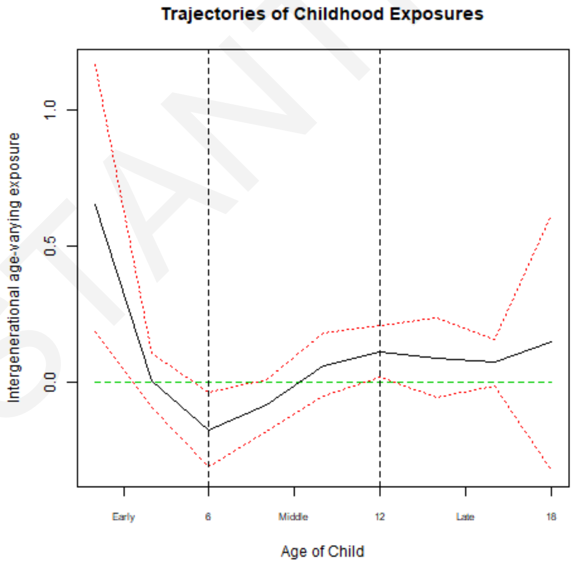
Figure 1.13: **Intergenerational Trajectories of Income by Family Structure**

This figure presents the results from model (1.5) based on the family structure and family shocks. Figures 1.13(a) and (b) display the estimates for intact and non-intact families, respectively. Figures 1.13(c) and (d) present the estimates for non-intact families when an early and a late family shock occurred, respectively.

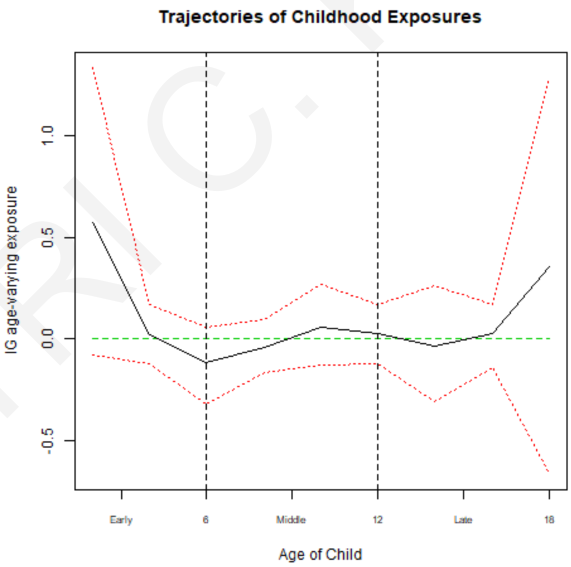
(a) Intact



(b) Non-Intact



(b) Non-Intact (early shock)



(c) Non-Intact (late shock)

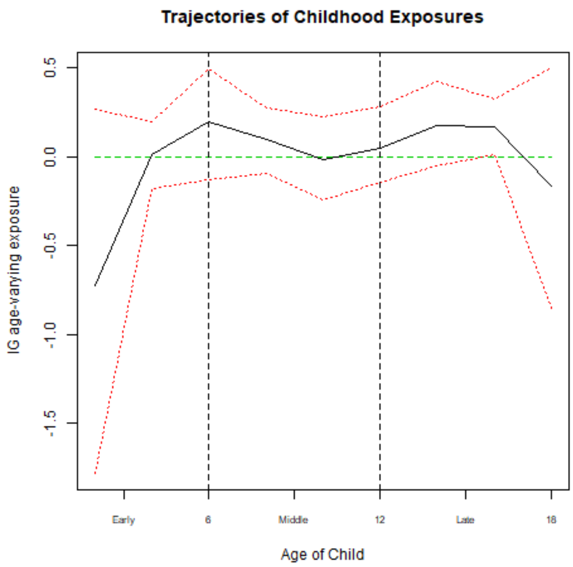
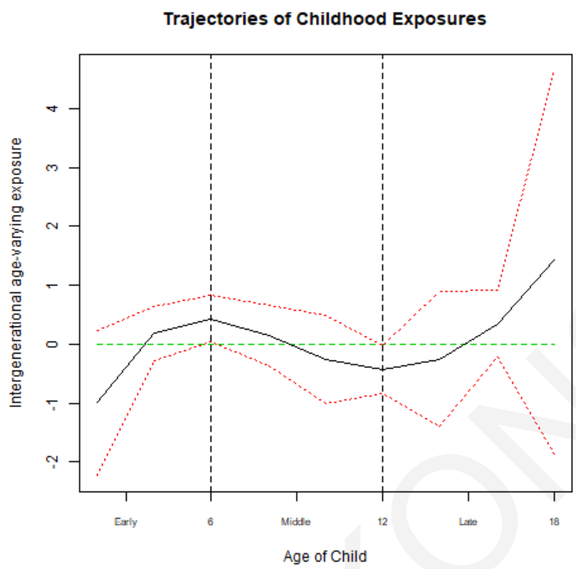


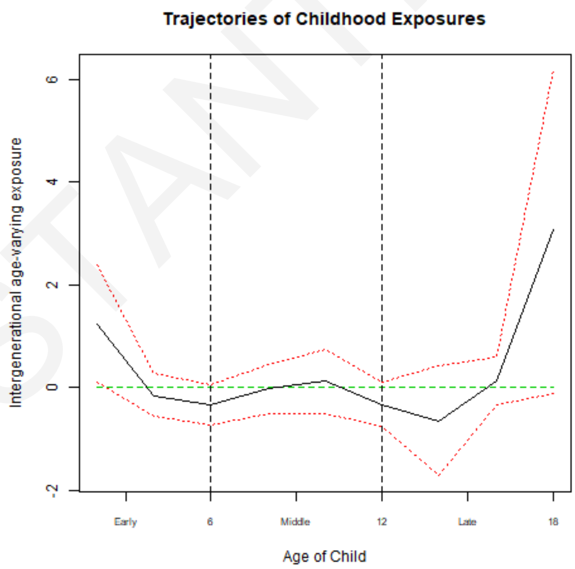
Figure 1.14: **Intergenerational Trajectories of the Stock of Income by Family Structure**

This figure presents the results from model (1.10) based on the family structure and family shocks. Figures 1.14(a) and (b) display the estimates for intact and non-intact families, respectively. Figures 1.14(c) and (d) present the estimates for non-intact families when an early and a late family shock occurred, respectively.

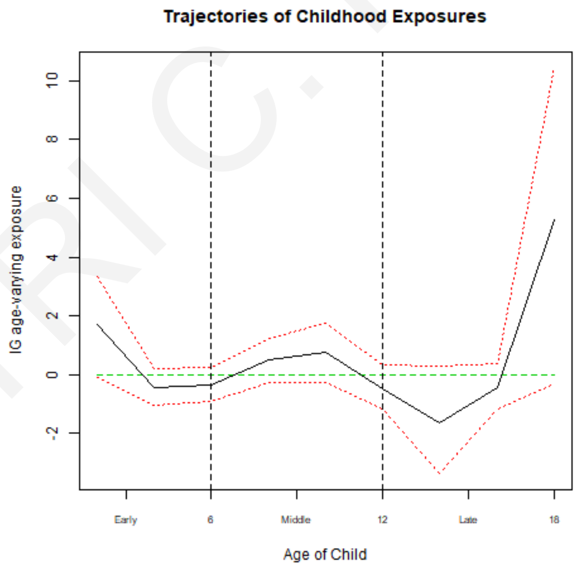
(a) Intact



(b) Non-Intact



(b) Non-Intact (early shock)



(c) Non-Intact (late shock)

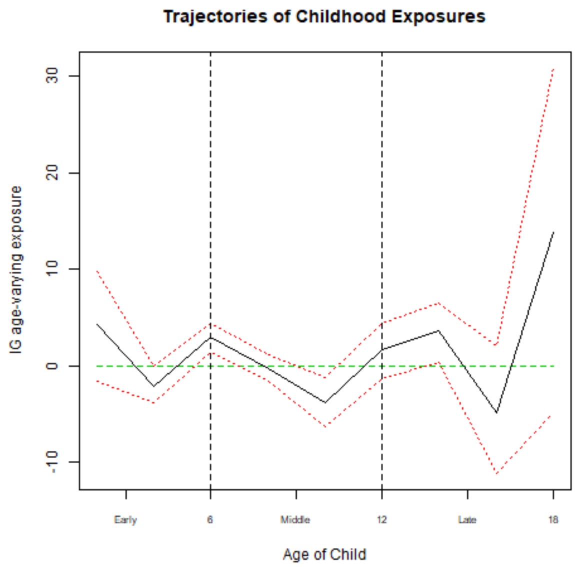
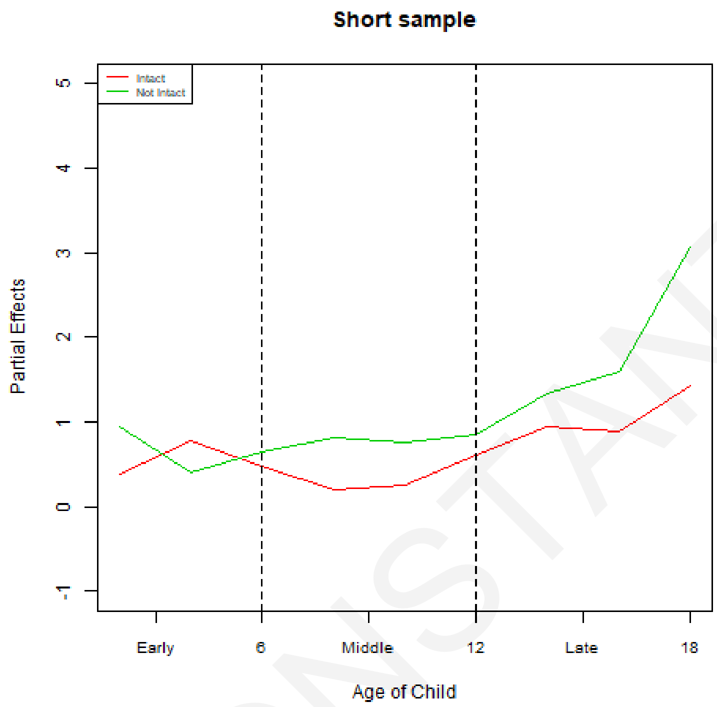


Figure 1.15: **Partial Effects of the Stock of Income by Family Structure**

(a) Intact vs Non-intact



(a) Non-intact (early) vs. Non-intact (late)

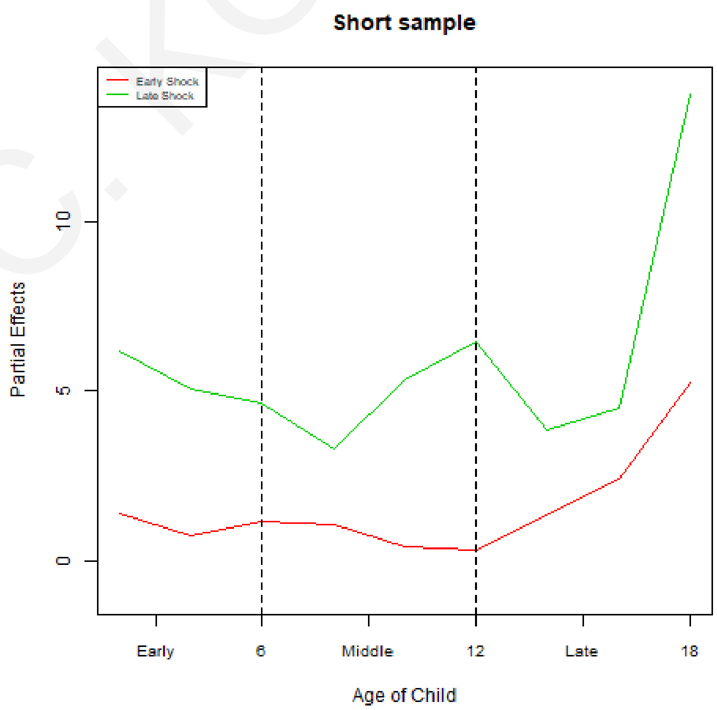
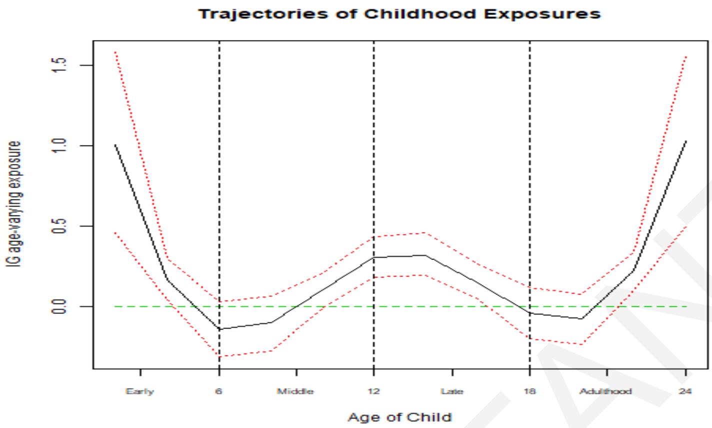
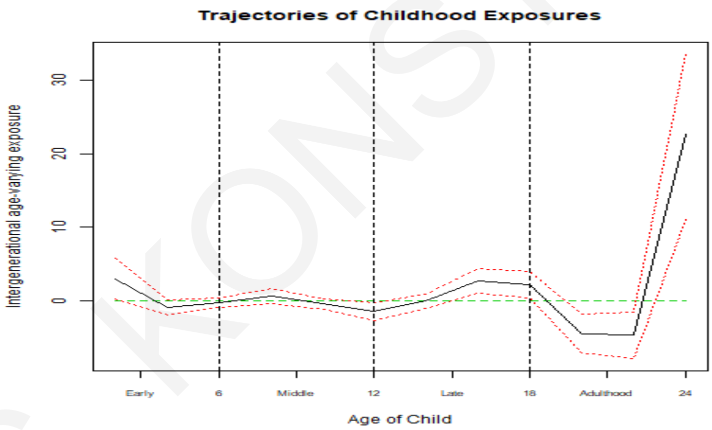


Figure 1.16: **Schooling Attainment: Intergenerational Trajectories**

(a) Bi-annual Income



(b) Stock of Income



(c) Partial Effects of Stock of Income

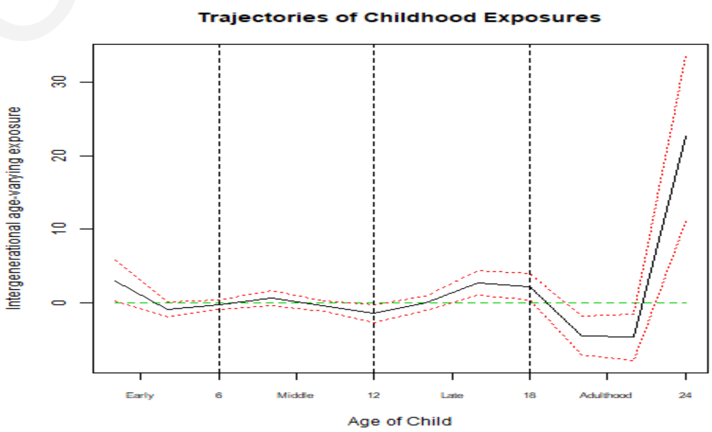
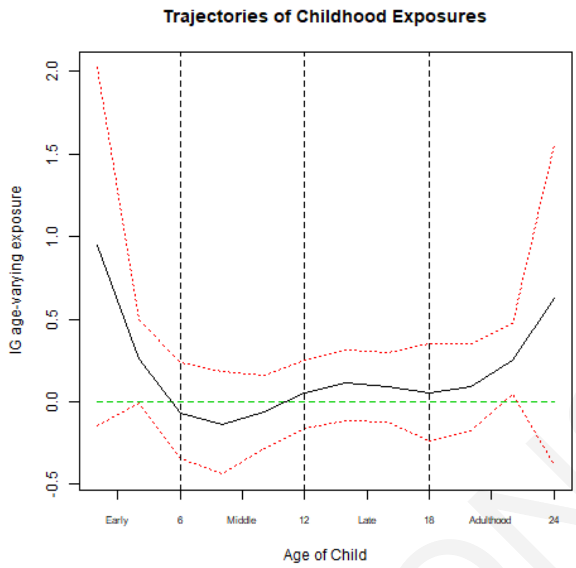
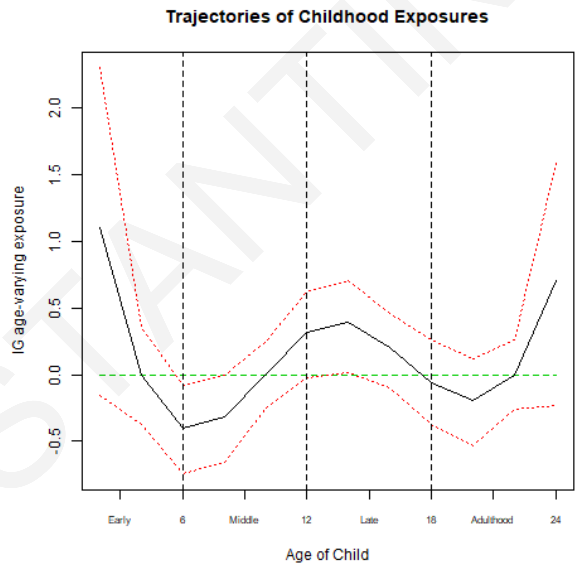


Figure 1.17: **Schooling Attainment: Intergenerational Trajectories of Income by Parental Income Quartiles**

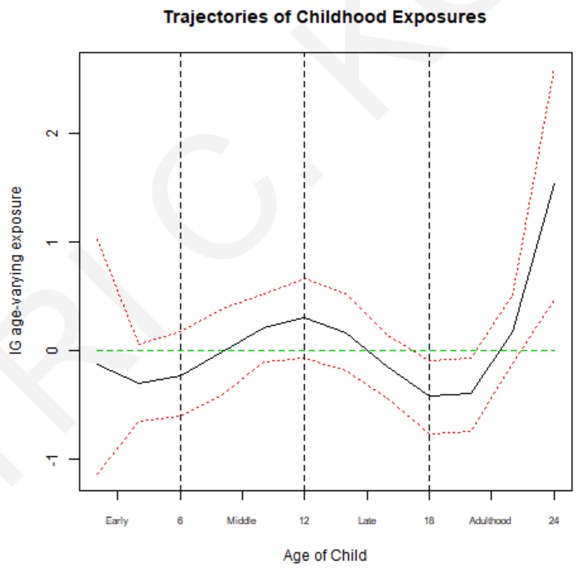
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

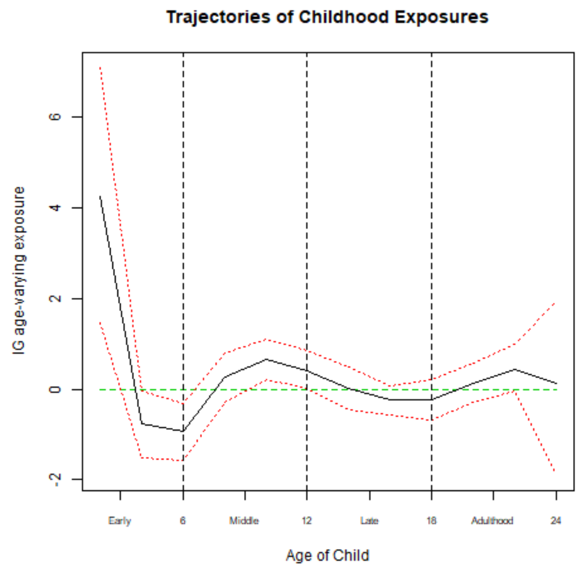
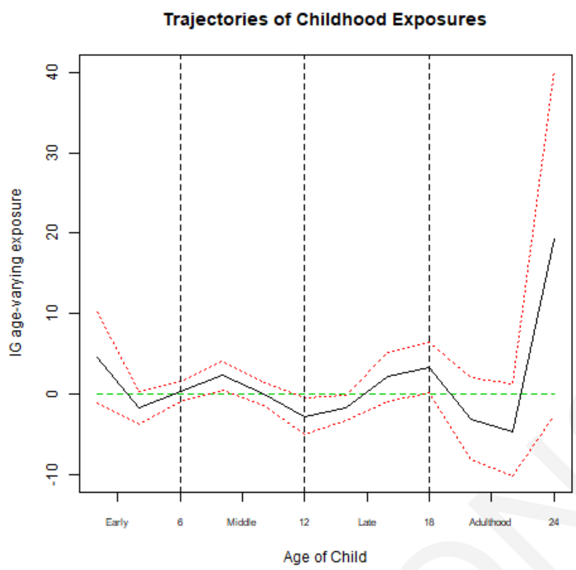
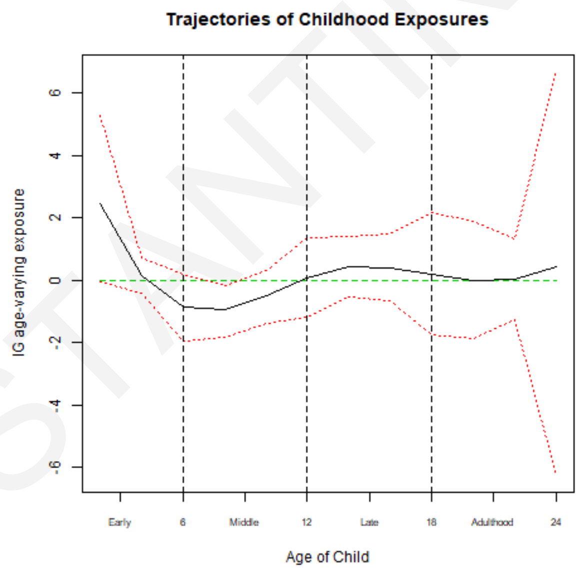


Figure 1.18: **Schooling Attainment: Intergenerational Trajectories of the Stock of Income by Parental Income Quartiles**

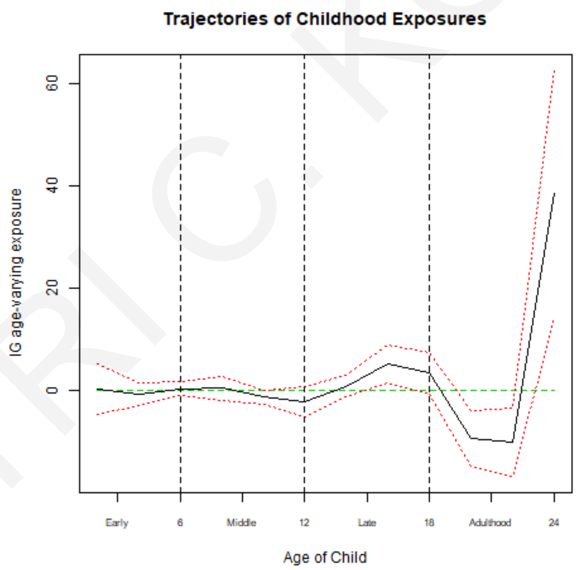
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

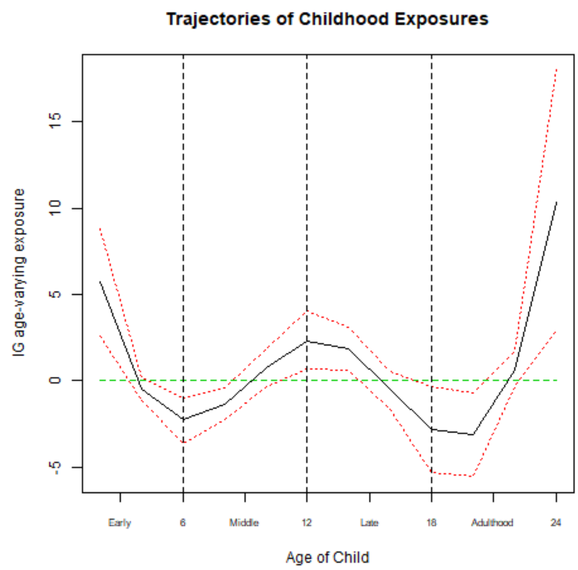


Figure 1.19: **Schooling Attainment: Partial Effects of the Stock of Income based by Parental Income Quartiles**

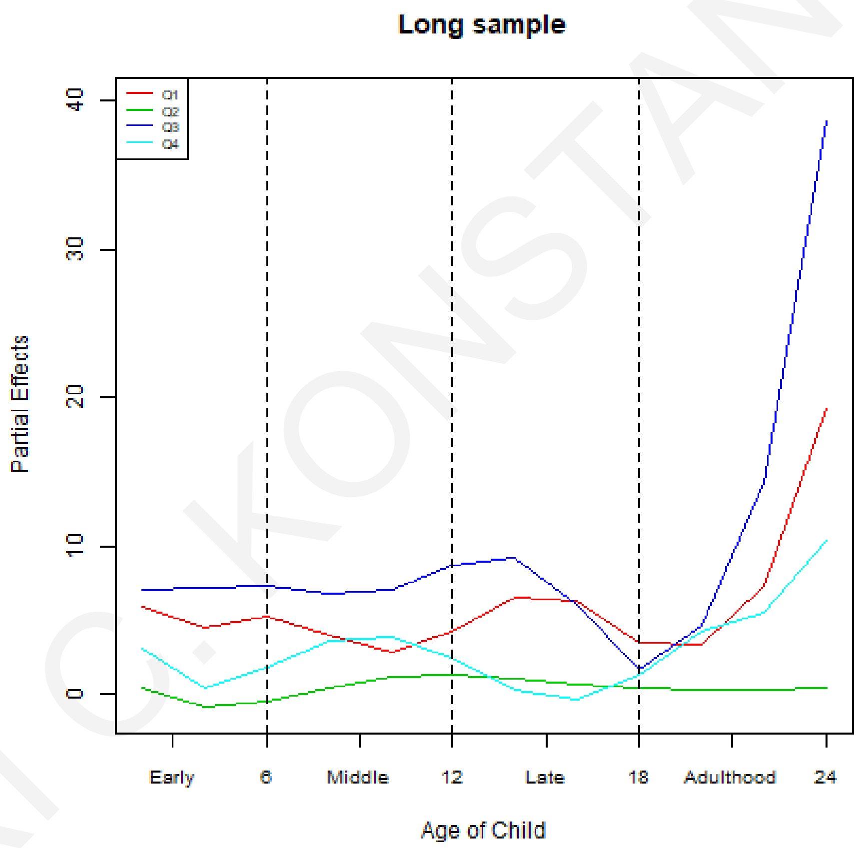
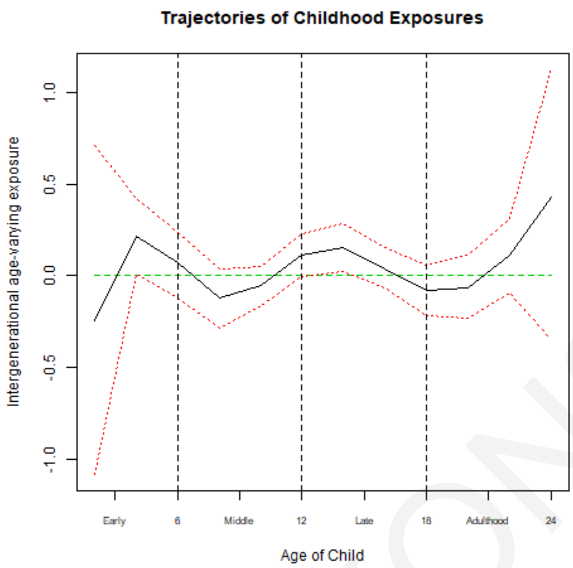


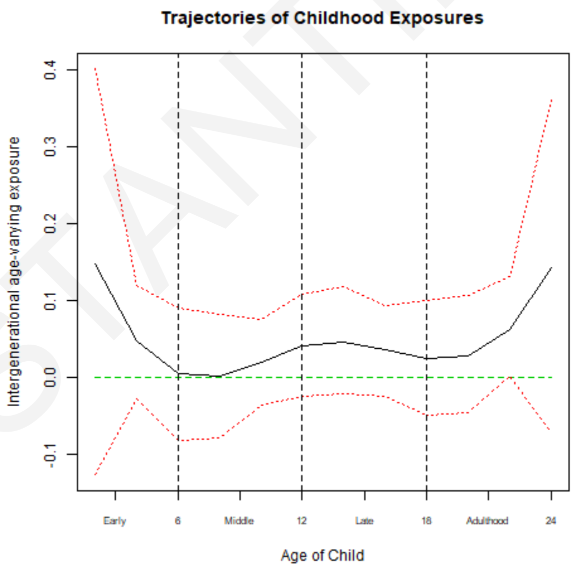
Figure 1.20: **Intergenerational Trajectories of Income for Daughters and Sons**

This figure presents the corresponding baseline results from models (1.5) and (1.10) for females and males. The red dotted lines represent 90% bootstrap confidence bands.

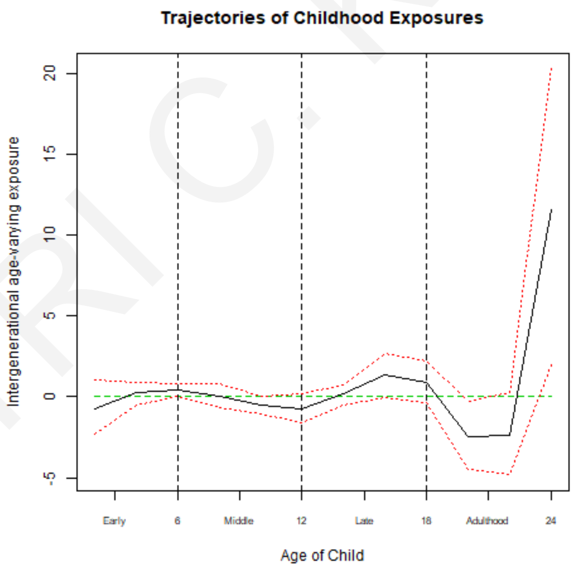
(a) Income - Daughters



(b) Income - Sons



(c) Stock of Income - Daughters



(d) Stock of Income - Sons

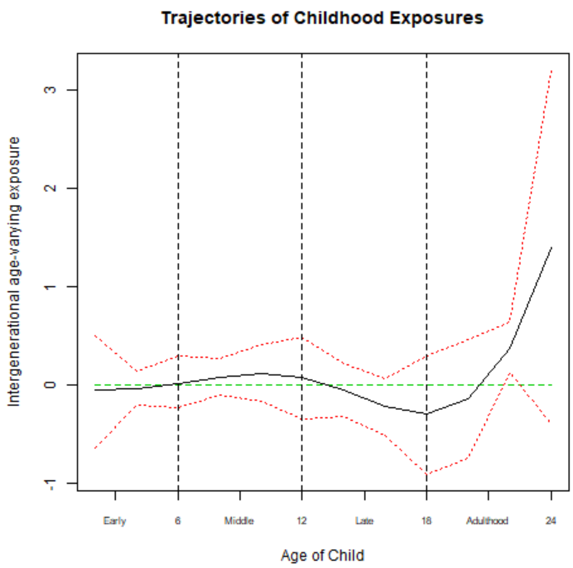


Figure 1.21: **Intergenerational Trajectories of Growth rates**

This figure presents the trajectory of income growth experiences.

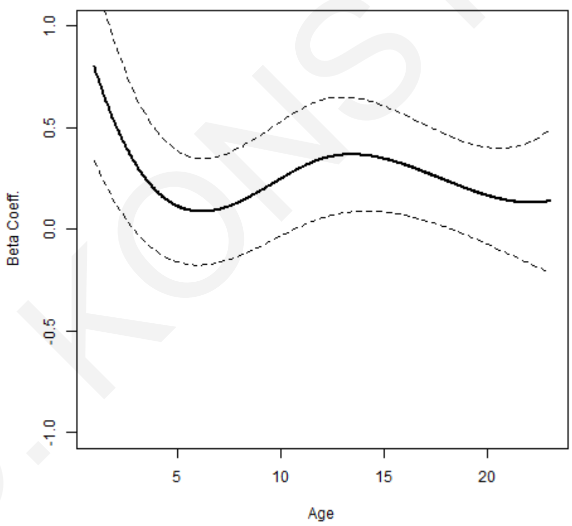
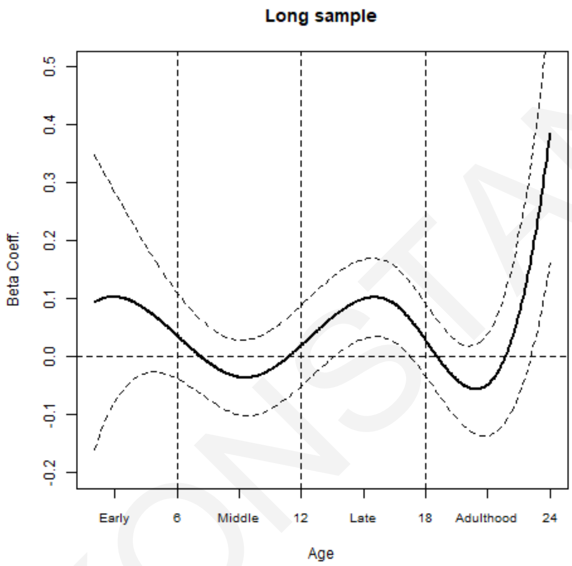


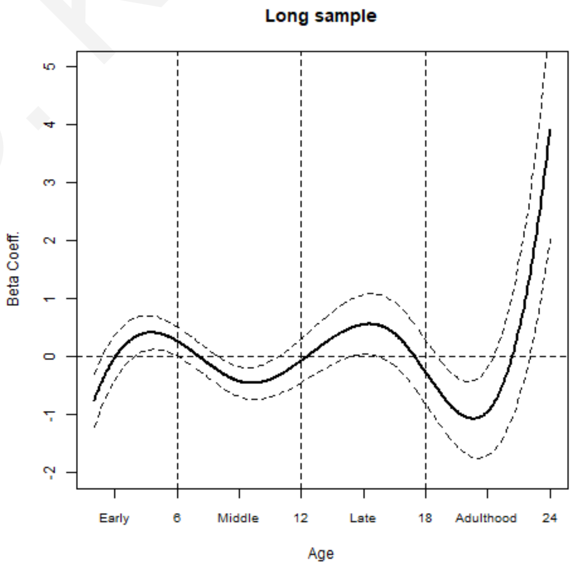
Figure 1.22: **Intergenerational Trajectories of Stock of Income by Father's Education**

This figure presents the baseline results from model (1.5) and (1.10) that include parent's education for annual income and stock of income and for both short and long samples.

(b) Biannual Income - Long Sample



(d) Stock of Income - Long Sample



1.9 Tables

Table 1.1: **Summary Statistics**

	Panel A. Baseline Sample			
	Short Sample ($n = 580$)			
	Mean	Std. Dev.	Min	Max
Labor Income of Individual	43649	22680	2646	105550
Parental Permanent Income	71438	23502	20417	158692
Age of mother at birth	25	5	16	44
Age of father at birth	28	6	17	48
	Long Sample ($n = 580$)			
	Mean	Std. Dev.	Min	Max
Labor Income of Individual	43538	21798	2646	105086
Parental Permanent Income	73997	24383	19649	158811
Age of mother at birth	25	5	16	46
Age of father at birth	28	6	17	48
	Panel B. Father's Education Sample			
	Short Sample ($n = 573$)			
	Mean	Std. Dev.	Min	Max
Labor Income of Individual	43815	22731	2646	105550
Parental Permanent Income	71805	23365	20417	158692
Father's Education	13	2	3	17
	Long Sample ($n = 677$)			
	Mean	Std. Dev.	Min	Max
Labor Income of Individual	43686	21842	2646	105086
Parental Permanent Income	26177	9078	7968	60557
Father's Education	13	2	3	17
	Panel C. Schooling Sample			
	Short Sample ($n = 820$)			
	Mean	Std. Dev.	Min	Max
Individual's Years of Schooling	14	2	8	17
Parental Permanent Income	26538	9223	7658	63234
	Long Sample ($n = 1392$)			
	Mean	Std. Dev.	Min	Max
Individual's Years of Schooling	14	2	8	17
Parental Permanent Income	26432	9683	5150	65437

Table 1.2: **Testing the Standard IGE Model**

This table presents the p-values for the LM test testing the null hypothesis of no omitted variable bias using both the short and long sample and two kinds of instruments. Columns 2 and 3 use as instrumental variables the annual parental income variables at age j one-at-a-time while Columns 4 and 5 use as instrumental variables the cumulative income up to age j.

	Annual		Stock	
Age	Short	Long	Short	Long
1	0.94	0.78	0.94	0.78
2	0.41	0.77	0.55	0.67
3	0.96	0.51	0.58	0.81
4	0.14	0.40	0.53	0.84
5	0.81	0.62	0.72	0.97
6	0.75	0.90	0.52	0.83
7	0.06	0.05	0.37	0.34
8	0.02	0.02	0.05	0.04
9	0.04	0.09	0.06	0.05
10	0.10	0.10	0.02	0.04
11	0.46	0.23	0.03	0.02
12	0.85	0.93	0.13	0.09
13	0.86	0.46	0.03	0.02
14	0.40	0.89	0.03	0.04
15	0.03	0.25	0.07	0.02
16	0.02	0.30	0.68	0.04
17	0.91	0.25	0.01	0.01
18	0.01	0.36	0.09	0.08
19	-	0.11	-	0.55
20	-	0.44	-	0.81
21	-	0.16	-	0.57
22	-	0.23	-	0.92
23	-	0.50	-	0.92
24	-	0.31	-	0.78

Chapter 2

Threshold Spatial Autoregression

2.1 Introduction

The study of social influences on individual behavior has attracted a lot of interest in economics recently. The idea of social interactions is that individual choices are directly influenced by the characteristics and choices of others. When choices are driven by social factors there exists complementarity among agents in a group (e.g., classroom, neighborhood) that generates interdependencies. This means that there exist incentives for an individual to behave similarly to others either because of social norms, social identity, peer effects, etc. The surveys by Durlauf and Ioannides (2010) and Benhabib, Bisin, and Jackson (2011a,b) discuss the various classes of social interaction models and their empirical applications.

The standard empirical models of social interactions are the linear-in-means model by Manski (1993) and the spatial autoregression mixed regression (e.g., Anselin (1988)). One problem with this type of models is that the linear functional form rules out interesting phenomena.[1] For example, consider an idealistic intergenerational model of poverty traps where the equilibrium law-of-motion is described by an intergenerational dynamic relationship between the child's permanent income and parent's permanent income conditional on the permanent income of other individuals in the neighborhood and whether parent face credit constraints when making human capital decision about their child. Such a model can be captured by a simple generalization of the linear

[1]Since the work of Schelling social interactions models have been used to explain a range of phenomena including growth (Cooper and John (1988)), technology adoption (Durlauf (1993)), crime (Glaeser, Sacerdote, and Scheinkman (1996)), delinquent behavior (Card and Giuliano (2013), occupations (Pan (2015)) among others. A common characteristic of all these models is their ability to generate multiple equilibria, which implies that a small change in fundamentals may lead to large differences in group behavior. That is, there exists a tipping point in the strength of social interactions that leads to an abrupt emergence of two distinct regimes in the underlying outcome variable.

mixed regressive spatial autoregressive model that allows for poverty traps depending on whether parents' income is above or below a threshold value γ.

$$y^c_{i,n} = \kappa_2 + \delta_\kappa I(y^p_{i,n} \leq \gamma) + \alpha \sum_{j \neq i} w_{ij,n} y^c_{j,n} + \beta y^p_{i,n} + e_{i,n}, \quad (2.1)$$

where $y^c_{i,n}$ is the child's log permanent income and $y^p_{i,n}$ is the parents' income. $w_{ij,n}$ are weights that capture the degree of influence that individual j has on i and define her social network. $\delta_\kappa = \kappa_1 - \kappa_2$ and the indicator function, $I(y^p_{i,n} \leq \gamma)$ takes the value 1 if $y^p_{i,n} \leq \gamma$ and 0 otherwise, one. When β, which is interpreted as the intergenerational elasticity (IGE) between the parent and child, is close to zero, parents' income is a weaker predictor of child's income implying greater mobility. In contrast, when β is close to one, the child's position in the income distribution is more dependent on her background. Note that when $\delta_\kappa = 0$ and $w_{ij,n} = 1$ we obtain the linear-in-means model. It is straightforward to see that the linear model-in-means model or the linear mixed regressive spatial autoregressive model will be biased because the omitted threshold term $\delta_\alpha I(y^p_{i,n} \leq \gamma)$, is correlated with parents' income. Hence, the estimated IGE will not consistently estimate the structural parameter β. More generally, the threshold effects may be present in the other parameters beyond the intercept which may exacerbate the bias. Therefore, in this paper we generalize the above simple model to allow general threshold type nonlinearities including threshold effects in the spatial autoregressive coefficient α, network weights $w_{ij,n}$, and regression coefficients β.

In particular, this paper develops a new class of social interaction models that generalize the spatial autoregressive model to allow for threshold type nonlinearities by proposing a general Threshold Spatial Autoregressive (TSAR) Model, which nests both mixed regressive, spatial autoregressive model as well as the spatial autoregressive model. Our model allows for regime specific endogenous and as well as contextual effects. Endogenous effects occur when the tendency of an agent's behavior depends on the group behavior while the choices are simultaneously determined. Exogenous or contextual effects occur when agent's behavior depends on the characteristics of others in the group. Endogenous effects are captured by spatial lags in our framework and constitute the most salient effects as their presence is often associated with social multipliers, multiple equilibria, and phase transitions. Specifically, we allow for either the network to be different across regimes or the marginal rate of substitution between private and social components of utility to be different across regimes or both.

Our estimation method generalizes the GMM estimation method of Lee (2007) to the case of threshold regression and develop a statistical theory for the threshold parameter as well as the regression coefficients including the spatial autoregressive coefficients. Specifically, we consider a two-step GMM estimator under the assumption of independent but heteroskedastic errors. The first-step estimates the objective function of GMM estimator for each value of the threshold parameter in a bounded set of values using an initial weight matrix. The estimated threshold parameter is then obtained as the argument that minimizes this objective function. The spatial autoregression coefficients and the slope coefficients are computed given the estimated threshold parameter. Finally, in the second-step, we obtain a quasi-optimal GMM estimator using a regime-specific weighting matrix by constructing appropriate moment functions using the first-step GMM estimator.

We contribute to the literature by providing a unifying framework in which we can subsume most commonly used social interaction models while allowing for general threshold effects. Our model can be viewed as a generalization of the Spatial Autoregressive Model (SAR) - mixed regressive model to allow for threshold type nonlinearities. The SAR mixed regressive model was proposed by Cliff and Ord (1973) and its 2SLS/IV estimation was studied by Anselin (1980), Kelejian and Prucha (1998), and Lee (2007) among others. GMM estimation was first studied by Lee (2007) and further studied by Lin and Lee (2010) among others. Interestingly, recently, there is an interest in nonlinear spatial regression models (e.g., Su and Jin (2010), Malikov and Sun (2017)).[2] However, none of these studies consider threshold type nonlinearities. Inference in threshold regressions is generally difficult. Chan (1993) showed that the asymptotic distribution of the threshold estimator depends on many nuisance parameters including the marginal distribution of the regressors. To overcome this difficulty, Hansen (2000) assumed that the difference between the slope coefficients of the two regimes decreases as the sample size grows and derived a useful asymptotic approximation, albeit nonstandard. Our framework nests both the fixed and diminishing threshold. As in Seo and Shin (2016) we exploit the smoothness of the GMM criterion to show consistency and asymptotic normality for our estimators.

The rest of the paper is organized as follows. In Section 2.2, we propose a general TSAR model. Section 2.3 presents our GMM estimation method. Section 2.4 derives limiting results for the proposed estimators and Section 2.5 proposes bootstrap inference. Section 2.6 reports Monte Carlo simulation results to assess the finite sample performance of our methods. Section 2.7 discusses future work and 2.8 concludes. Finally, we delay all the mathematical proofs in the Appendix. We define the column and row sum matrix norms of an $n \times n$ matrix A as $||A||_1$ and $||A||_\infty$ respectively, and the spectral norm $||A||_{sp} = \lambda_{max}^{1/2}(AA')$.

[2] There is also a literature that considers social interaction models in discrete choice models (e.g., Brock and Durlauf (2001a, 2002)), duration models (e.g., Sirakaya (2006), de Paula (2009)), and regression discontinuity methods (RDD) (e.g., Card and Rothstein (2008)).

2.2 TSAR Model

We propose a new class of social interaction models that generalize the Spatial Autoregression - Mixed Regression to allow for threshold effects in the endogenous as well as contextual effects.

Let the ith individual's outcome be denoted by $y_{i,n}$ and her individual characteristics by a k_1-dimensional vector $x_{i,n}$ of stochastic variables, where the first element of this vector is a constant. We consider an $n \times n$ fixed (predetermined) weight matrix W_n which has zero diagonal elements and its (i,j)th element is denoted by $w_{ij,n}$. We split our sample based on a threshold random variable $z_{i,n}$. Specifically we consider the following threshold model with social interactions

$$y_{i,n} = \begin{cases} \alpha_{11} y_{i,n}^{m-}(w,\lambda) + \alpha_{12} y_{i,n}^{m+}(w,\lambda) + \beta_1' x_{i,n} + \gamma_{11}' x_{i,n}^{m-}(w,\lambda) + \gamma_{12}' x_{i,n}^{m+}(w,\lambda) + e_{i,n}, & z_{i,n} \leq \lambda \\ \alpha_{21} y_{i,n}^{m-}(w,\lambda) + \alpha_{22} y_{i,n}^{m+}(w,\lambda) + \beta_2' x_{i,n} + \gamma_{21}' x_{i,n}^{m-}(w,\lambda) + \gamma_{22}' x_{i,n}^{m+}(w,\lambda) + e_{i,n}, & z_{i,n} > \lambda \end{cases} \tag{2.2}$$

where $\{e_{i,n}\}$ is a sequence of independent errors with zero mean and finite variance, and we denote $y_{i,n}^m(w) = \sum_{j \neq i} w_{ij,n} y_{j,n}$, $x_{i,n}^m(w) = \sum_{j \neq i} w_{ij,n} x_{j,n}$, $w_{ij,n}^-(\lambda) = w_{ij,n} 1\{z_{j,n} \leq \lambda\}$, $y_{i,n}^{m-}(w,\lambda) = \sum_{j \neq i} w_{ij,n}^-(\lambda) y_{j,n}$, $x_{i,n}^{m-}(w,\lambda) = \sum_{j \neq i} w_{ij,n}^-(\lambda) x_{j,n}$, $y_{i,n}^{m+}(w,\lambda) = y_{i,n}^m(w) - y_{i,n}^{m-}(w,\lambda)$, and $x_{i,n}^{m+}(w,\lambda) = x_{i,n}^m(w) - x_{i,n}^{m-}(w,\lambda)$.

Our model includes two kinds of endogenous social effects that capture social pressures. Let us focus on the lower regime; the first term (α_{11}) captures regime specific social effects while the second term (α_{12}) captures across regime social effects. In general we would expect the former term to capture conformity effects while the second term to capture nonconformity effects.

Rearranging model (2.2) yields

$$\begin{aligned} y_{i,n} &= \alpha y_{i,n}^m(w) + \delta_{\alpha_2} y_{i,n}^{m-}(w,\lambda) + \beta' x_{i,n} + \gamma' x_{i,n}^m(w) + \delta_{\gamma_2}' x_{i,n}^{m-}(w,\lambda) \\ &+ \delta_{\alpha_1} y_{i,n}^m(w) 1\{z_{i,n} \leq \lambda\} + \delta_{\alpha\alpha} y_{i,n}^{m-}(w,\lambda) 1\{z_{i,n} \leq \lambda\} + \delta_\beta' x_{i,n} 1\{z_{i,n} \leq \lambda\} \\ &+ \delta_\gamma' x_{i,n}^m(w) 1\{z_{i,n} \leq \lambda\} + \delta_{\gamma\gamma}' x_{i,n}^{m-}(w,\lambda) 1\{z_{i,n} \leq \lambda\} + e_{i,n} \end{aligned} \tag{2.3}$$

where $\alpha = \alpha_{22}$, $\beta = \beta_2$, $\gamma = \gamma_{22}$, $\delta_{\alpha_2} = \alpha_{21} - \alpha_{22}$, $\delta_{\gamma_2} = \gamma_{21} - \gamma_{22}$, $\delta_{\alpha_1} = \alpha_{12} - \alpha_{22}$, $\delta_\beta = \beta_1 - \beta_2$, $\delta_{\alpha\alpha} = (\alpha_{11} - \alpha_{12}) - (\alpha_{21} - \alpha_{22})$, $\delta_\gamma = \gamma_{12} - \gamma_{22}$ and $\delta_{\gamma\gamma} = (\gamma_{11} - \gamma_{12}) - (\gamma_{21} - \gamma_{22})$.

Equation (2.3) nests several theoretically appealing models. When modeling social effects in the context of threshold models it is reasonable to assume that social effects are formed based only on regime-specific information, that is, $\alpha_{12} = \alpha_{21} = \gamma_{12} = \gamma_{21} = 0$. In contrast, when $\delta_{\alpha_2} = \delta_{\gamma_2} = \delta_{\alpha\alpha} = \delta_{\gamma\gamma} = 0$ then we obtain a social interaction model that assumes that the formation of social effects occurs regardless of the regime; it is solely based on the reference group. Finally, equation (2.3) nests the mixed spatial autoregressive (MRSAR) model when $\delta_{\alpha_2} = \gamma = \delta_{\gamma_2} = \delta_{\alpha_1} = \delta_{\alpha\alpha} = \delta_\beta = \delta_\gamma = \delta_{\gamma\gamma} = 0$.

Let $x_i(\lambda) = x_{i,n}1\{z_{i,n} \leq \lambda\}$ and define the matrix $X_n(\lambda)$ and X_n by stacking the elements $x_i(\lambda)$ and x_i, respectively. And, we introduce three $n \times n$ weight matrices $W_n(\lambda)$, $W^-_{n,\lambda}$ and $W^-_n(\lambda)$ whose (i,j)th element equals $w^-_{ij,n}(\lambda) = w_{ij,n}1\{z_{j,n} \leq \lambda\}$, $w^{*-}_{ij,n}(\lambda) = w_{ij,n}1\{z_{i,n} \leq \lambda\}$ and $w^-_{\lambda,ij,n} = 1\{z_{i,n} \leq \lambda\}w^-_{ij,n}(\lambda)$, respectively. We then rewrite model (2.3) in matrix form

$$Y_n = \alpha_0 W_n Y_n + X_n(W_n)\theta_{\beta_0} + Y_n(W_n, \lambda_0)\delta_{\alpha_0} + X_n(W_n, \lambda_0)\delta_{\theta_{\beta_0}} + e_n \tag{2.4}$$

where $Y_n = [y_{1,n}, \ldots, y_{n,n}]'$, $e_n = [e_{1,n}, \ldots, e_{n,n}]'$.

Also, we denote $X_n(W_n) = [X_n, W_n X_n]$, $Y_n(W_n, \lambda) = [W^-_{n,\lambda}Y_n, W_n(\lambda)Y_n, W^-_n(\lambda)Y_n]$, $X_n(W_n, \lambda) = [X_n(\lambda), W_n(\lambda)X_n,\ W^-_{n,\lambda}X_n, W^-_n(\lambda)X_n]$, $\theta_\beta = (\beta', \gamma')'$ to be a $(2k_1) \times 1$ vector of slope coefficients of the upper regime, and $\delta_\alpha = (\delta_{\alpha_1}, \delta_{\alpha_2}, \delta_{\alpha\alpha})'$ and $\delta_{\theta_\beta} = (\delta'_\beta, \delta'_{\gamma_2}, \delta'_\gamma, \delta'_{\gamma\gamma})'$ to be the 3×1 and $(4k_1) \times 1$ vectors of threshold effects. In (2.4), $\alpha_0, \theta_{\beta_0}$, δ_{α_0} and $\delta_{\theta_{\beta_0}}$ denote the true parameters of the model. This model nests the linear social interactions model when $\delta_{\alpha_0} = \delta_{\theta_{\beta_0}} = 0$, under which model (2.4) is the so-called spatial Durbin regression model.

Let $S_n(\theta_y, \lambda) = I_n - \alpha W_n - \delta_{\alpha_1} W^-_{n,\lambda} - \delta_{\alpha_2} W_n(\lambda) - \delta_{\alpha\alpha} W^-_n(\lambda)$, where $\theta_y = (\alpha, \delta'_\alpha)'$ and the errors for any possible values of the parameter space $e_n(\theta) = S_n(\theta_y, \lambda)Y_n - X_n(W_n)\theta_\beta - X_n(W_n, \lambda)\delta_{\theta_\beta}$, where $\theta = (\alpha, \delta'_\alpha, \theta'_\beta, \delta'_{\theta_\beta}, \lambda)'$. Moreover, we denote $\theta^* = (\theta'_\beta, \delta'_{\theta_\beta})'$ and $\theta^{**} = (\theta'_\beta, \delta'_{\theta_\beta}, \alpha, \delta'_\alpha)'$. Then the reduced form is given by

$$Y_n = S_n^{-1} X_n(W_n)\theta_{\beta_0} + S_n^{-1} X_n(W_n, \lambda_0)\delta_{\theta_{\beta_0}} + S_n^{-1} e_n, \tag{2.5}$$

if S_n is non-singular, where $S_n = S_n(\theta_{y0}, \lambda_0)$. This generally leads to
$E\left[(A_n Y_n)' e_n\right] = \mathrm{tr}\left[(A_n S_n^{-1})' E\left(e_n e_n'\right)\right] \neq 0$ for $A_n = W_n$, W^-_{n,λ_0}, $W_n(\lambda_0)$, and $W^-_n(\lambda_0)$ where $\mathrm{tr}(\cdot)$ is the trace operator.

We start with regularity conditions on the weight matrices, W_n, $W^-_{n,\lambda}$, $W_n(\lambda)$, $W^-_n(\lambda)$ and $S_n(\theta_y, \lambda)$.

Assumption 1.

(1.1) $\rho\left(\alpha_0 W_n + \delta_{\alpha_1,0} W^-_{n,\lambda_0} + \delta_{\alpha_2,0} W_n(\lambda_0) + \delta_{\alpha\alpha,0} W^-_n(\lambda_0)\right) < 1$, where $\rho(A)$ is the largest eigenvalue of matrix A in absolute values, and the spatial weight matrix W_n and S_n^{-1} have finite row- and column-sum norm.

(1.2) The parameter vector $\theta_0 = (\alpha_0, \delta'_{\alpha_0}, \theta'_{\beta_0}, \delta'_{\theta_{\beta_0}}, \lambda_0)'$ is an interior point of a compact set $\Theta = \Theta^{**} \times \Lambda$ in the Euclidean Space R^{k_θ}, where $k_\theta = 6k_1 + 5$ and $\Lambda = \left[\underline{\lambda}, \bar{\lambda}\right]$. Also, $\delta_{\alpha_{1,0}} = c_{\alpha_1} n^{-a}$, $\delta_{\alpha_{2,0}} = c_{\alpha_2} n^{-a}$, $\delta_{\alpha\alpha,0} = c_{\alpha\alpha} n^{-a}$, $\delta_{\beta_0} = c_\beta n^{-a}$, $\delta_{\gamma_0} = c_\gamma n^{-a}$, $\delta_{\gamma_{2,0}} = c_{\gamma_2} n^{-a}$ and $\delta_{\gamma\gamma,0} = c_{\gamma\gamma} n^{-a}$ with $c_{\alpha_1}, c_{\alpha_2}, c_{\alpha\alpha}, c_\beta, c_\gamma, c_{\gamma_2}, c_{\gamma\gamma} \neq 0$ and $0 \leq a < \frac{1}{2}$.

Assumption (1.1) is a crucial assumption for the analysis of spatial estimators, as it imposes limits on the correlations of the spatial elements. Following Seo and Shin (2016) our framework nests both the fixed threshold effect as in Chan (1993) and diminishing threshold effect framework of Hansen (2000). Assumption (1.2) states that the threshold effect gets small as the sample size increases, when $0 < a < \frac{1}{2}$, while when $a = 0$, we have the fixed threshold effect.

We proceed by stating the assumptions necessary to derive the asymptotic properties of our estimator

Assumption 2.

(2.1) $\{e_{i,n}\}$'s are independent $(0, \sigma_i^2)$ errors, independent of $\{z_{i,n}\}$ with finite moments larger than the fourth order. Moreover, $\max_{1\leq i\leq n} E|e_{i,n}|^{4+\eta} < M < \infty$ for some $\eta > 0$.

(2.2) The threshold variable $z_{i,n}$ is i.i.d. with a continuous and bounded density, $f(\cdot)$, such that $f(\lambda_0) > 0$.

(2.3) We consider a linear transformation of the moment equations $a_n g_n(\theta)$, where a_n is a matrix with a full row rank greater than or equal to k_θ and converges to a constant full row rank matrix a_0.

This set of assumptions is similar to Seo and Shin (2016). Assumptions (2.1)-(2.2) are also used in Hansen (2000). Assumption(2.3) is crucial for deriving the asymptotic distribution of our GMM estimator.

As opposed to Seo and Shin (2016), we use both linear and quadratic moments as suggested by Lee (2007). However, our moment conditions are different than those of Lee (2007) due to the existence of regime specific social effects and the fact that the threshold parameter is unknown and needs to be estimated.

Let Q_n be an $n \times k_Q$ matrix of initial instrumental variables (IV) with $k_Q > k_{\theta^{**}}$. For example, we can use X_n, $W_n X_n$, $W_n^2 X_n$, $W_n^3 X_n$,... after removing linearly dependent components.[3] The moment conditions corresponding to the orthogonality conditions of Q_n and e_n are $E(Q_n' e_n) = 0$. The quadratic moments are based on a class of constant $n \times n$ matrices denoted by $\mathcal{P}_{1n} = \{P_n : \text{diag}(P_n) = 0\}$. As shown in Lin and Lee (2010), by selecting matrices from $\mathcal{P}_{1n}$, the corresponding quadratic moments are defined as follows: $E[(P_{jn} e_n)' e_n] = 0$, where matrices P_{jn} are selected from $\mathcal{P}_{1n}$, for $j = 1, \dots, m$.

[3]While the initial matrix of instruments is not regime specific, we propose regime specific instruments that allow us to obtain a more efficient second-step estimator.

Assumption 3. The matrices in $\mathcal{P}_{1n}$ are uniformly bounded in both row and column sums in absolute values. The elements of Q_n are uniformly bounded. Both X_n and Q_n have full rank.

The GMM estimator is derived from the minimization of $J_n(\theta) = g_n(\theta)' a_n' a_n g_n(\theta)$. The set of moment conditions $E(g_n(\theta))$ for the GMM estimation is given by a $(k_Q + m) \times 1$ vector where $g_n(\theta)$ is the set of empirical moments

$$g_n(\theta) = \begin{bmatrix} e_n(\theta)' P_{1n} e_n(\theta) \\ \vdots \\ e_n(\theta)' P_{mn} e_n(\theta) \\ Q_n' e_n(\theta) \end{bmatrix}. \tag{2.6}$$

Before proceeding with the identification condition we introduce the following notation Define $G_n = W_n S_n^{-1}$, $G_n(\lambda) = W_n(\lambda) S_n^{-1}$, $G_{n,\lambda}^{-} = W_{n,\lambda}^{-} S_n^{-1}$ and $G_n^{-}(\lambda) = W_n^{-}(\lambda) S_n^{-1}$. Additionally, $X_{n,\lambda}^{*}$ =$[X_n(W_n), X_n(W_n, \lambda)]$ and $X_n^{*} = X_{n,\lambda_0}^{*}$, $\theta^{*} = (\theta_\beta', \delta_{\theta_\beta}')'$ and $\theta_0^{*} = (\theta_{\beta_0}', \delta_{\theta_{\beta_0}}')'$.
Furthermore,
$\tilde{X}(W_n, \lambda) = (X_n^{*}, X_n^{*} - X_{n,\lambda}^{*}, G_n X_n^{*} \theta_0^{*}, G_n(\lambda_0) X_n^{*} \theta_0^{*}, G_{n,\lambda_0}^{-} X_n^{*} \theta_0^{*}, G_n^{-}(\lambda_0) X_n^{*} \theta_0^{*},$
$(G_n(\lambda_0) - G_n(\lambda)) X_n^{*} \theta_0^{*}, (G_{n,\lambda_0}^{-} - G_{n,\lambda}^{-}) X_n^{*} \theta_0^{*}, (G_n^{-}(\lambda_0) - G_n^{-}(\lambda)) X_n^{*} \theta_0^{*}),$
$\tilde{\Theta} = ((\theta_0^{*} - \theta^{*})', \theta^{*\prime}, \alpha_0 - \alpha, \delta_{\alpha_{2,0}} - \delta_{\alpha_2}, \delta_{\alpha_{1,0}} - \delta_{\alpha_1}, \delta_{\alpha\alpha,0} - \delta_{\alpha\alpha}, \delta_{\alpha_2}, \delta_{\alpha_1}, \delta_{\alpha\alpha})$.

Then the error term can be decomposed as follows

$$e_n(\theta) = d_n(\theta) + \left[A_n(\theta_y, \lambda) + I_n\right] e_n \tag{2.7}$$

where $d_n(\theta) = \tilde{X}(W_n, \lambda)\tilde{\Theta}'$ and
$A_n(\theta_y, \lambda) = (\alpha_0 - \alpha) G_n + (\delta_{\alpha_{2,0}} - \delta_{\alpha_2}) G_n(\lambda_0) + (\delta_{\alpha_{1,0}} - \delta_{\alpha_1}) G_{n,\lambda_0}^{-} + (\delta_{\alpha\alpha,0} - \delta_{\alpha\alpha}) G_n^{-}(\lambda_0)$
$+ \delta_{\alpha_2}(G_n(\lambda_0) - G_n(\lambda)) + \delta_{\alpha_1}(G_{n,\lambda_0}^{-} - G_{n,\lambda}^{-}) + \delta_{\alpha\alpha}(G_n^{-}(\lambda_0) - G_n^{-}(\lambda))$.

From equation (2.7), we can see that while the first part of $e_n(\theta)$ depends on all parameters of the model, the second term only depends only on the spatial autoregressive parameters θ_y and the threshold parameter λ.

Taking the expectation of (2.6) and using the above notation we obtain

$$E(g_n(\theta)) = \begin{bmatrix} E(d_n(\theta)' P_{1n} d_n(\theta)) + \operatorname{tr}\{\Gamma_n E(A_n(\theta_y, \lambda)' P_{1n}^{s})\} + \operatorname{tr}\{\Gamma_n E(A_n(\theta_y, \lambda)' P_{1n} A_n(\theta_y, \lambda)\} \\ \vdots \\ E(d_n(\theta)' P_{mn} d_n(\theta)) + \operatorname{tr}\{\Gamma_n E(A_n(\theta_y, \lambda)' P_{mn}^{s})\} + \operatorname{tr}\{\Gamma_n E(A_n(\theta_y, \lambda)' P_{mn} A_n(\theta_y, \lambda)\} \\ E(Q_n' d_n(\theta)) \end{bmatrix}$$

where $P^s_{jn} = P_{jn} + P'_{jn}$ for $j = 1, 2, \ldots, m$ and $\Gamma_n = diag(\sigma_1^2, \ldots, \sigma_n^2)$. The parameter θ_0 is identified as long as $a_0 \lim_{n\to\infty} n^{-1} E(g_n(\theta)) = 0$ has a unique root at $\theta_0 \in \Theta$, where a_0 is a constant full rank matrix. We then examine under which conditions linear and quadratic moments have a unique solution at θ_0 and the identification condition is summarized in Assumption 4.

The linear moments corresponding to $E(Q'_n d_n(\theta))$ will have a unique solution under Assumption (4.1), where

$$G(Q_n, \theta) = \frac{\partial E(Q'_n d_n(\theta))}{\partial \theta'} = \left[\; G_{\theta_y}(\lambda), \quad G_{\theta_\beta}, \quad G_{\delta_{\theta_\beta}}(\lambda), \quad G_\lambda(\delta_\alpha, \delta_{\theta_\beta}, \lambda) \;\right]$$

is a $k_Q \times k_\theta$ matrix with

$$\begin{aligned}
G_{\theta_y}(\lambda) &= \frac{\partial E(Q'_n d_n(\theta))}{\partial \theta'_y} = -E\left[Q'_n G_n X^*_n \theta^*_0, Q'_n G_n(\lambda) X^*_n \theta^*_0, Q'_n G^-_{n,\lambda} X^*_n \theta^*_0, Q'_n G^-_n(\lambda) X^*_n \theta^*_0\right] \\
G_{\theta_\beta} &= \frac{\partial E(Q'_n d_n(\theta))}{\partial \theta'_\beta} = -E\left[Q'_n X_n(W_n)\right] \\
G_{\delta_{\theta_\beta}}(\lambda) &= \frac{\partial E(Q'_n d_n(\theta))}{\partial \delta'_{\theta_\beta}} = -E\left[Q'_n X_n(W_n, \lambda)\right]
\end{aligned}$$

and

$$\begin{aligned}
&G_\lambda(\delta_\alpha, \delta_{\theta_\beta}, \lambda) = \frac{\partial E(Q'_n d_n(\theta))}{\partial \lambda} \\
= \; &-f(\lambda)\left[E(Q'_n X_n|\lambda), E(Q'_n W_n X_n|\lambda), E(Q'_n W_n X_n|\lambda), E(Q_n W_n(\lambda) X_n|\lambda) + E(Q_n W^-_{n,\lambda} X_n|\lambda)\right]\delta_{\theta_\beta} \\
&-f(\lambda)\left[E(Q'_n G_n X^*_n \theta^*_0|\lambda), E(Q'_n G_n X^*_n \theta^*_0|\lambda), E(Q_n G_n(\lambda) X^*_n \theta^*_0|\lambda) + E(Q_n G^-_{n,\lambda} X^*_n \theta^*_0|\lambda))\right]\delta_\alpha.
\end{aligned}$$

A trivial violation of the rank condition, Assumption (4.1), occurs if (i) $\theta^*_0 = 0$ or X_n is irrelevant, or (ii) $\delta_{\alpha_0} = 0$ and $\delta_{\theta_{\beta_0}} = 0$ or threshold effect does not exist.

From (2.5), we have

$$\begin{aligned}
Y_n &= S_n^{-1} X^*_n \theta^*_0 + u_n = (I_n - S_n) S_n^{-1} X^*_n \theta^*_0 + X^*_n \theta^*_0 + u_n \\
&= [G_n X^*_n \theta^*_0, G_n(\lambda_0) X^*_n \theta^*_0, G^-_{n,\lambda_0} X^*_n \theta^*_0, G^-_n(\lambda_0) X^*_n \theta^*_0]\theta_{y0} + X^*_n \theta^*_0 + u_n \qquad (2.8)
\end{aligned}$$

where $u_n = S_n^{-1} e_n$ and u_n therefore follows a SAR model, $u_n = (\alpha_0 W_n + \delta_{\alpha_{1,0}} W^-_{n,\lambda_0} + \delta_{\alpha_{2,0}} W_n(\lambda_0) + \delta_{\alpha\alpha,0} W^-_n(\lambda_0)) u_n + e_n$. Assuming that threshold effect does exist and X_n contain relevant regressors in predicting Y_n, we consider an example that the rank condition of Assumption (4.1) is violated. Specifically, in parallel to Lee (2007) we consider the following example that X^*_n and $[G_n X^*_n \theta^*_0, G_n(\lambda_0) X^*_n \theta^*_0, G^-_{n,\lambda_0} X^*_n \theta^*_0, G^-_n(\lambda_0) X^*_n \theta^*_0]$ are linearly dependent. That is, there exists a $6k_1 \times 4$ non-zero constant matrix c_0 such that $X^*_n c_0 = [G_n X^*_n \theta^*_0, G_n(\lambda_0) X^*_n \theta^*_0, G^-_{n,\lambda_0} X^*_n \theta^*_0, G^-_n(\lambda_0) X^*_n \theta^*_0]$. Then, model (2.8) becomes

$$Y_n = X^*_n(c_0 \theta_{y0} + \theta^*_0) + u_n, \qquad (2.9)$$

and $G_n(Q_n, \theta_0)$ does not have full rank as

$$\left[G_{\theta_y}(\lambda_0), G_{\theta_\beta}, G_{\delta_{\theta_\beta}}(\lambda_0)\right] = -E\left[Q_n' X_n^* c_0, Q_n' X_n(W_n), Q_n' X_n(W_n, \lambda_0)\right]$$

is linearly dependent in column. For model (2.9), if $\left(\theta_{y0}', \lambda_0\right)'$ is identified from using the quadratic moments, we have $0 = E\left[Q_n' d_n(\theta)\right] = E\left[Q_n' X_n^* (\theta_0^* - \theta^*)\right]$ so that θ_0^* is identified as X_n^* has full rank. That is, the model is identified under Assumption (4.2) if Assumption (4.1) fails to hold.

Assumption 4

(4.1) $G(Q_n, \theta)$ has a full rank over $\theta \in \Theta$ and $\max_{\lambda \in \Lambda} F(\lambda) < 1$; or

(4.2) (i) $E(Q_n' X_n^*)$ has the full rank $6k_1$ and is linearly independent of $n^a G_\lambda(\delta_{\alpha 0}, \delta_{\theta_{\beta 0}}, \lambda_0)$;
(ii) D_n has the full rank 4 for some $m \geq 4$, where we denote

$$D_n = \begin{bmatrix} \operatorname{tr}\left[\Gamma_n E\left(P_{1n}^s G_n\right)\right] & \dots & \operatorname{tr}\left[\Gamma_n E\left(P_{mn}^s G_n\right)\right] \\ \operatorname{tr}\left\{\Gamma_n E\left[P_{1n}^s G_n(\lambda_0)\right]\right\} & \dots & \operatorname{tr}\left\{\Gamma_n E\left[P_{mn}^s G_n(\lambda_0)\right]\right\} \\ \operatorname{tr}\left\{\Gamma_n E\left[P_{1n}^s G_{n,\lambda_0}^-\right]\right\} & \dots & \operatorname{tr}\left\{\Gamma_n E\left[P_{mn}^s G_{n,\lambda_0}^-\right]\right\} \\ \operatorname{tr}\left\{\Gamma_n E\left[P_{1n}^s G_n^-(\lambda_0)\right]\right\} & \dots & \operatorname{tr}\left(\Gamma_n E\left\{\Gamma_n E\left[P_{mn}^s G_n^-(\lambda_0)\right]\right\}\right) \end{bmatrix}.$$

Assumption (4.1) is a global identification condition, while Assumption (4.2) is a local identification condition derived from equation (B8), given in Appendix B, evaluated at the true parameter value. The global condition, when Assumption (4.1) fails to hold, has a lot less transparent expression than the local identification condition and requires extra notation defined for equation (B8).

2.3 Estimation

In this section, we describe our two-step GMM estimators, where the first-step GMM estimator is an initial estimator that is consistent and asymptotically normally distributed, but its asymptotic efficiency can be improved further as all the IVs and the P_{jn}'s matrices in the quadratic moment conditions can be non optimal according to the reduce form model (2.5).

Step 1. Given an initial weight matrix $a_n' a_n$, the GMM estimator of θ is given by $\hat{\theta} = \underset{\theta \in \Theta}{arg\ min}\ J_n(\theta)$, where $J_n(\theta) = g_n(\theta)' a_n' a_n g_n(\theta)$. However, for practical reasons we estimate the objective function of GMM estimator for each value of the threshold parameter in a bounded set of values $\Lambda = [\underline{\lambda}, \bar{\lambda}]$. Conditional on λ, we can obtain the GMM estimator of θ^{**} by concentration

$$\widehat{\theta}^{**}(\lambda) = \underset{\theta^{**}}{arg\ min} J_n(\theta^{**}, \lambda),$$

where $J_n(\theta^{**}, \lambda) = g_n(\theta^{**}, \lambda)' a_n' a_n g_n(\theta^{**}, \lambda)$. Then, we obtain the GMM estimator of θ by

$$\hat{\lambda} = \underset{\lambda \in \Lambda}{arg\ min}\ J_n(\hat{\theta}^{**}(\lambda), \lambda) \quad \text{and} \quad \widehat{\theta}^{**}(\hat{\lambda}) = \underset{\theta^{**}}{arg\ min} J_n(\theta^{**}, \hat{\lambda}) \tag{2.10}$$

Step 2. We improve the efficiency of our initial estimator by considering regime specific moment functions. Motivated by the reduce form model (2.5) we define the empirical moments

$$\hat{g}_n(\theta) = \begin{bmatrix} e_n(\theta)' \hat{P}_{1n} e_n(\theta) \\ \vdots \\ e_n(\theta)' \hat{P}_{4n} e_n(\theta) \\ \hat{Q}_n' e_n(\theta) \end{bmatrix} \tag{2.11}$$

where

$$\begin{aligned}
\hat{P}_{1n} &= W_n S_n^{-1}(\hat{\theta}_y, \hat{\lambda}) - \text{diag}\left(W_n S_n^{-1}(\hat{\theta}_y, \hat{\lambda})\right) \\
\hat{P}_{2n} &= W_{n,\hat{\lambda}}^{-} S_n^{-1}(\hat{\theta}_y, \hat{\lambda}) - \text{diag}\left(W_{n,\hat{\lambda}}^{-} S_n^{-1}(\hat{\theta}_y, \hat{\lambda})\right) \\
\hat{P}_{3n} &= W_n(\hat{\lambda}) S_n^{-1}(\hat{\theta}_y, \hat{\lambda}) - \text{diag}\left(W_n(\hat{\lambda}) S_n^{-1}(\hat{\theta}_y, \hat{\lambda})\right) \\
\hat{P}_{4n} &= W_n^{-}(\hat{\lambda}) S_n^{-1}(\hat{\theta}_y, \hat{\lambda}) - \text{diag}\left(W_n^{-}(\hat{\lambda}) S_n^{-1}(\hat{\theta}_y, \hat{\lambda})\right)
\end{aligned}$$

and

$$\begin{aligned}
\hat{Q}_n = \ & [W_n S_n^{-1}(\hat{\theta}_y, \hat{\lambda})[X_n(W_n)\hat{\theta}_\beta, X_n(W_n, \hat{\lambda})\hat{\delta}_{\theta_\beta}], \\
& W_{n,\hat{\lambda}}^{-} S_n^{-1}(\hat{\theta}_y, \hat{\lambda})[X_n(W_n)\hat{\theta}_\beta, X_n(W_n, \hat{\lambda})\hat{\delta}_{\theta_\beta}], \\
& W_n(\hat{\lambda}) S_n^{-1}(\hat{\theta}_y, \hat{\lambda})[X_n(W_n)\hat{\theta}_\beta, X_n(W_n, \hat{\lambda})\hat{\delta}_{\theta_\beta}], \\
& W_n^{-}(\hat{\lambda}) S_n^{-1}(\hat{\theta}_y, \hat{\lambda})[X_n(W_n)\hat{\theta}_\beta, X_n(W_n, \hat{\lambda})\hat{\delta}_{\theta_\beta}], \\
& X_n(W_n), X_n(W_n, \hat{\lambda})]
\end{aligned}$$

after removing redundant terms.

In addition, given the initial consistent estimator calculated in the first step, we can construct a matrix $\hat{\Omega}_n$ to estimate $\Omega_n = E[g_n(\theta_0) g_n(\theta_0)']$, where $g_n(\theta)$ equals $\hat{g}_n(\theta)$ with $\hat{\theta}$ replaced with θ_0 and the mathematical expression of Ω_n as shown in Theorem

1 in the next section with $m = 4$, we define

$$\hat{\Omega}_n = \begin{bmatrix} \mathrm{tr}\left(\hat{\Gamma}_n \hat{P}_{1n}\left(\hat{\Gamma}_n \hat{P}_{1n}\right)^s\right) & \dots & \mathrm{tr}\left(\hat{\Gamma}_n \hat{P}_{1n}\left(\hat{\Gamma}_n \hat{P}_{4n}\right)^s\right) & 0'_{k_Q} \\ \vdots & \ddots & \vdots & \vdots \\ \mathrm{tr}\left(\hat{\Gamma}_n \hat{P}_{4n}\left(\hat{\Gamma}_n \hat{P}_{1n}\right)^s\right) & \dots & \mathrm{tr}\left(\hat{\Gamma}_n \hat{P}_{4n}\left(\hat{\Gamma}_n \hat{P}_{4n}\right)^s\right) & 0'_{k_Q} \\ 0_{k_Q} & \dots & 0_{k_Q} & \hat{Q}'_n \hat{\Gamma}_n \hat{Q}_n \end{bmatrix} \tag{2.12}$$

where $\hat{\Gamma}_n = diag\left\{\hat{e}^2_{1,n}, \dots, \hat{e}^2_{n,n}\right\}$ is an $n \times n$ diagonal matrix and $\hat{e}_i$ is the ith element of the estimated residual $\hat{e}_n = S_n(\hat{\theta}_y, \hat{\lambda})Y_n - X_n(W_n)\hat{\theta}_\beta - X_n(W_n, \hat{\lambda})\hat{\delta}_{\theta_\beta}$. Then, the second-step GMM estimator of θ is given by

$$\tilde{\theta} = \underset{\theta \in \Theta}{arg\ min}\ \ \hat{g}_n(\theta)'\hat{\Omega}_n^{-1}\hat{g}_n(\theta) \tag{2.13}$$

but evaluated through concentration of the objective function in a similar way as in the first step.[4]

2.4 Asymptotic Theory

In this section we develop a statistical theory for the threshold parameter as well as the regression coefficients including the spatial autoregressive coefficients. Proposition 1 shows the consistency of the GMM estimator, while Theorem 1 shows that the asymptotic distribution of the GMM estimator follows a normal distribution. The following assumption imposes a pointwise law of large numbers result that is used to show the consistency of our estimator.

Assumption 5. For any given $\theta \in \Theta$, (i) $n^{-1}Q'_n\chi_n X^*_n\theta^*_0 = E\left(Q'_n\chi_n X^*_n\theta^*_0\right) + o_p(1)$, where $\chi_n = G_n$, $G_n(\lambda)$, $G^-_{n,\lambda}$, and $G^-_n(\lambda)$; (ii) $n^{-1}Q'_n X^*_{n,\lambda}\theta^* = n^{-1}E\left(Q'_n X^*_{n,\lambda}\theta^*\right) + o_p(1)$; (iii) $n^{-1}d'_n(\theta)\left(\sum_{j=1}^m a_{nj}P_{jn}\right)d_n(\theta) = n^{-1}E\left[d'_n(\theta)\left(\sum_{j=1}^m a_{nj}P_{jn}\right)d_n(\theta)\right] + o_p(1)$.

Assumption 5 is a high-end assumption for the law of large numbers, where Assumption 5(i) ensures $n^{-1}Q'_n d_n(\theta) = n^{-1}E\left(Q'_n d_n(\theta)\right) + o_p(1)$ holds for $\theta \in \Theta$, and Assumption 5(ii) is a law of large number result for a second-order U statistic of $d_n(\theta) = X^*_n\theta^*_0 + (\alpha_0 - \alpha)G_n X^*_n\theta^*_0 - \delta_{\alpha_2}G_n(\lambda)X^*_n\theta^*_0 - \delta_{\alpha_1}G^-_{n,\lambda}X^*_n\theta^*_0 - \delta_{\alpha\alpha}G^-_n(\lambda)X^*_n\theta^*_0 - X^*_{n,\lambda}\theta^*$. It is beyond the current scope of this paper to give a rigid proof of such LLN results, we therefore assign it as an assumption.

Proposition 2.1 *Under Assumptions 1-5, the identification condition holds which im-*

[4] A third-step estimator can further provide efficiency gains using residuals and parameter estimates from second-step.

plies that $a_0 \lim_{n\to\infty} n^{-1}E(g_n(\theta)) = 0$ *has a unique root at* $\theta_0 \in \Theta$*, and the GMM estimator is consistent.*

Theorem 1 Under Assumptions 1-5, we have

$$\sqrt{n}H_n^{-1}\begin{pmatrix} \hat{\theta}^{**}-\theta_0^{**} \\ \hat{\lambda}-\lambda_0 \end{pmatrix} = \begin{bmatrix} \sqrt{n}\left(\hat{\theta}^{**}-\theta_0^{**}\right) \\ n^{\frac{1}{2}-a}\left(\hat{\lambda}-\lambda_0\right) \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0},\Sigma_n\right), \tag{2.14}$$

where we denote $H_n = diag\left(I_{6k_1+4}, n^a\right)$,

$$\Sigma_n = \lim_{n\to\infty}\left[\frac{1}{n}H_n\Lambda_n' a_n' a_n \frac{1}{n}\Lambda_n H_n\right]^{-1}\left[\frac{1}{n}H_n\Lambda_n' a_n' a_n \frac{1}{n}\Omega_n a_n' a_n \frac{1}{n}\Lambda_n H_n\right]\left[\frac{1}{n}H_n\Lambda_n' a_n' a_n \frac{1}{n}\Lambda_n H_n\right]^{-1},$$

$$\begin{aligned}\Omega_n &= E\left[g_n(\theta_0)g_n(\theta_0)'\right] \\ &= \begin{bmatrix} \text{tr}\left(\Gamma_n P_{1n}\left(\Gamma_n P_{1n}\right)^s\right) & \dots & \text{tr}\left(\Gamma_n P_{1n}\left(\Gamma_n P_{mn}\right)^s\right) & 0_{k_Q}' \\ \vdots & \ddots & \vdots & \vdots \\ \text{tr}\left(\Gamma_n P_{mn}\left(\Gamma_n P_{1n}\right)^s\right) & \dots & \text{tr}\left(\Gamma_n P_{mn}\left(\Gamma_n P_{mn}\right)^s\right) & 0_{k_Q}' \\ 0_{k_Q} & \dots & 0_{k_Q} & E(Q_n'\Gamma_n Q_n) \end{bmatrix},\end{aligned} \tag{2.15}$$

and $\Lambda_n = -\partial E(g_n(\theta_0))/\partial\theta'$ and $\Lambda_n H_n$ has full column rank.

Applying Lemma A.1 in Lin and Lee (2010) gives (2.15). And, by (B8) in Appendix B, we have

$$\Lambda_n = \begin{bmatrix} \text{tr}\left[\Gamma_n E\left(P_{1n}^s G_n\right)\right] & \text{tr}\left\{\Gamma_n E\left[P_{1n}^s G_n(\lambda_0)\right]\right\} & \text{tr}\left\{\Gamma_n E\left[P_{1n}^s G_{n,\lambda_0}^-\right]\right\} & \text{tr}\left\{\Gamma_n E\left[P_{1n}^s G_n^-(\lambda_0)\right]\right\} & 0_{6k_1}' & \text{tr}\left\{\Gamma_n E\left[P_{1n}^s \varphi_2\left(\theta_{y,0},\lambda_0\right)\right]\right\} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{tr}\left[\Gamma_n E\left(P_{mn}^s G_n\right)\right] & \text{tr}\left\{\Gamma_n E\left[P_{mn}^s G_n(\lambda_0)\right]\right\} & \text{tr}\left\{\Gamma_n E\left[P_{mn}^s G_{n,\lambda_0}^-\right]\right\} & \text{tr}\left\{\Gamma_n E\left[P_{mn}^s G_n^-(\lambda_0)\right]\right\} & 0_{6k_1}' & \text{tr}\left\{\Gamma_n E\left[P_{mn}^s \varphi_2\left(\theta_{y,0},\lambda_0\right)\right]\right\} \\ E(Q_n' G_n X_n^*\theta_0^*) & E(Q_n' G_n(\lambda_0) X_n^*\theta_0^*) & E(Q_n' G_{n,\lambda_0}^- X_n^*\theta_0^*) & E(Q_n' G_n^-(\lambda_0) X_n^*\theta_0^*) & E(Q_n' X_n^*) & G_\lambda(\delta_{\alpha_0},\delta_{\theta_{\beta_0}},\lambda_0) \end{bmatrix}. \tag{2.16}$$

Our model nests both Lee (2007) and Seo and Shin (2016). Equation (3.2) of Lee (2007) is a submatrix of Λ_n as it does not contain the regime specific components. On the other hand, excluding the spatial matrices will allow us obtain Seo and Shin (2016) model. This theorem says that our estimator follows the normal distribution asymptotically regardless of whether $a = 0$ or $0 < a < \frac{1}{2}$, since we exploit the smoothness of the GMM criterion to reduce the rate of convergence as in Seo and Shin (2016). For the estimation of the threshold parameter, λ_0, the GMM estimator converges at slower speed than what one would expect from a least-square based estimator, see Chan (1993) and Hansen (2000) for example. However, the GMM estimator of λ_0 enjoys the conveience of constructing the classic t-statistic in making inference of $\lambda = \lambda_0$.

While the GMM estimator in Theorem 1 allows for an arbitrary unknown heteroskedasticity, it is not efficient. Our second step estimator aims at addressing this issue by obtaining a quasi-optimal estimator. In particular, by the generalized Schwartz inequality, the "optimal" weighting matrix is the inverse of the variance matrix $\Omega_n = E\left[g_n(\theta_0)g_n(\theta_0)'\right]$. In the case of homoskedastic errors, the best selection of P_{jn} and Q_n is available and thus, an optimal GMM is feasible. However, as argued by Lin and Lee (2010), in the heteroskedastic case, an optimal estimator may not be feasible because the best selection involves matrix Γ_n, which is generally unknown. In this sense, the second-step estimator is quasi-optimal and using a consistent estimator of Ω_n^{-1} we obtain a feasible quasi-optimal GMM estimator. The following proposition is used to support the use of $\hat{\Omega}_n$ in (2.13).

Proposition 2.2 *Under Assumptions 1-5, $n^{-1}(\hat{\Omega}_n - \Omega_n) = o_p(1)$, where $\hat{\Omega}_n$ is defined in (2.12) and Ω_n equals $\hat{\Omega}_n$ with $\hat{\theta}$ replaced with θ_0.*

Then, the quasi-optimal GMM estimator is derived from minimizing $\hat{g}_n(\theta)'\hat{\Omega}_n^{-1}\hat{g}_n(\theta)$ as long as $n^{-1}\hat{\Omega}_n$ is consistently estimated. Theorem 2 establishes that the quasi-optimal estimator will be asymptotically normal with variance $(\lim_{n\to\infty} n^{-1}H_n'\Lambda_n'\Omega_n^{-1}\Lambda_n H_n)^{-1}$ assuming the following regularity condition for Ω_n in Assumption 6. Finally, we obtain asymptotically valid inferences from the quasi-optimal GMM estimator by re-estimating Ω_n and Λ_n using the second-step residuals.

The variance matrix Ω_n must satisfy the following regularity condition.

Assumption 6. The matrix, $\lim_{n\to\infty} n^{-1}\Omega_n$, exists and is nonsingular.

Theorem 2 Under Assumptions 1-6, we obtain the quasi-optimal GMM estimator from minimizing $\hat{g}_n(\theta)'\hat{\Omega}_n^{-1}\hat{g}_n(\theta)$, which has the limiting distribution

$$\begin{bmatrix} \sqrt{n}\left(\tilde{\theta}^{**} - \theta_0^{**}\right) \\ n^{\frac{1}{2}-a}\left(\tilde{\lambda} - \lambda_0\right) \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0}, \Sigma_n^*\right), \tag{2.17}$$

with $\Sigma_n^* = (\lim_{n\to\infty} n^{-1}H_n\Lambda_n'\Omega_n^{-1}\Lambda_n H_n)^{-1}$, where Λ_n with $m = 4$ and P_{jn} equals $\hat{P}_{jn}$ with $\hat{\theta}$ replaced with θ_0 for $j = 1, 2, 3, 4$.

To make statistical inference on $\hat{\theta}$, we need to find a consistent estimator for Σ_n^*. However, $E\left[P_{jn}^s\varphi_2\left(\theta_{y,0},\lambda_0\right)\right]$ as an element of Λ_n contains terms like $f\left(\lambda_0\right)$ and $E\left(P_{jn}^s G_n|\lambda_0\right)$, $E\left(P_{jn}^s G_{n,\lambda_0}^-|\lambda_0\right)$, and $E\left(P_{jn}^s G_n\left(\lambda_0\right)|\lambda_0\right)$, and these unknown terms have to be estimated nonparametrically by kernel method, which makes the estimate of Λ_n depend on a bandwidth parameter value. The overall accuracy of an estimate of Λ_n may not be robust to the choice of the bandwidth parameter. Therefore, bootstrap-based inference (e.g., Anselin (1988) and Taspinar and Vijverberg (2018), Gupta (2018)) for $\hat{\theta}$, although time consuming, is recommended.

Next, we propose a bootstrap procedure for inference including a threshold test for linearity.

2.5 Bootstrap Inference

The bootstrap procedure for confidence intervals is summarized below

(i) Using the second-step estimator $\tilde{\theta}$ defined in equation (2.13), we obtain the demeaned residuals $\hat{\varepsilon}_n = \tilde{e}_n(\tilde{\theta}) - \iota' \tilde{e}_n(\tilde{\theta})/n$ where $\tilde{e}_n(\tilde{\theta}) = S_n(\tilde{\theta}_y, \tilde{\lambda})Y_n - X_n\left(W_n\right)\tilde{\theta}_\beta - X_n\left(W_n, \tilde{\lambda}\right)\tilde{\delta}_{\theta_\beta}$ and ι is the $n \times 1$ vector.

(ii) The bootstrap sample of residuals ε^*_n=$(\varepsilon^*_{1,n}, ..., \varepsilon^*_{n,n})'$ is generated from a two point distribution $P\left(\varepsilon^*_{i,n} = \hat{\varepsilon}_{i,n}(1+\sqrt{5})/2)\right)$=$(\sqrt{5}-1)/(2\sqrt{5})$ and $P\left(\varepsilon^*_{i,n} = \hat{\varepsilon}_{i,n}(1-\sqrt{5})/2)\right)$ $= (\sqrt{5}+1)/(2\sqrt{5})$, where $\hat{\varepsilon}_{i,n}$ is the ith demeaned residual. Then, these residuals are used to generate $y^*_n = S_n(\tilde{\theta}_y, \tilde{\lambda})^{-1}\left(X_n\left(W_n\right)\tilde{\theta}_\beta + X_n\left(W_n, \tilde{\lambda}\right)\tilde{\delta}_{\theta_\beta} + \varepsilon^*_n\right)$.

(iii) Using the bootstrap sample $\{y^*_{in}, x_{i,n}\}$, i=$1, .., n$ and weight matrix W_n and applying the estimation in Section 2.3 we obtain bootstrap analogs of the empirical moments $g^*_{b,n}(\theta)$ and the second-step GMM estimator $\tilde{\theta}^*_b$.

(iv) Repeat steps $(ii) - (iii)$ B times and compute bootstrap standard errors $s^*(\theta) = \sqrt{\hat{V}^*_n}$, where $\hat{V}^*_n = \frac{1}{B}\sum_{b=1}^{B}(\tilde{\theta}^*_b - \bar{\tilde{\theta}}^*_b)^2$ and $100(1-\alpha)\%$ bootstrap confidence intervals $[\tilde{\theta}_j - q^*(1-\alpha/2), \tilde{\theta}_j - q^*(\alpha/2)]$, where $\tilde{\theta}_j$ is an element j of the vector $\tilde{\theta}$ and $q^*(\cdot)$ is the quantile function of $\tilde{\theta}^*_b - \tilde{\theta}_b$.

Next, we consider testing the null hypothesis $H_0 : \delta_\alpha = \delta_{\theta_\beta} = 0$ under which the model (2.4) reduces to the linear spatial Durbin regression model (SAR). However, the inference is not standard because the threshold parameter λ is not identified under the null of the linear SAR model; this issue is also known as the Davies problem. Following Hansen (1996) we employ a bootstrap sup-LR type *supD* test that can be used to approximate its asymptotic distribution

$$supD = \sup_{\lambda}(J_{SAR} - J_{TSAR}(\lambda)). \tag{2.18}$$

where $J_{TSAR}(\lambda) = \hat{g}_n(\tilde{\theta}(\lambda))'\hat{\Omega}_n^{-1}\hat{g}_n(\tilde{\theta}(\lambda))$ where $\tilde{\theta}(\lambda)$ is the second-step estimator for a given λ and $J_{SAR} = \hat{g}_n(\breve{\theta})'\hat{\Omega}_n^{-1}\hat{g}_n(\breve{\theta}))$ where $\breve{\theta}$ is the second-step GMM estimator under the restrictions $\delta_\alpha = \delta_\theta = 0$, $\tilde{\theta}$ is the estimator (2.13) and $\hat{g}_n$ are the empirical moments defined in (3.9).

The bootstrap residuals u^*_i, $i = 1, ..., n$ are generated with random draws from $i.i.d.$ $N(0,1)$. Then, set $y^*_i = \breve{e}_i u^*_i$, where $\breve{e}_i = \tilde{e}_i - \bar{\tilde{e}}_i$. For each bootstrap sample $\{y^*_{b,in}, x_{i,n}\}$,

i=1, .., n and weight matrix W_n, we estimate the second step GMM estimator of the SAR model and obtain its J-statistic J^b_{SAR}, from the b^{th} bootstrap sample. For each $\lambda \in \Lambda$, estimate the second step GMM estimator of the TSAR model and obtain its J-stat $J^b_{TSAR}(\lambda)$. Next, construct the Distance statistic $supD^*_b = \sup_\lambda(J^b_{SAR} - J^b_{TSAR}(\lambda))$. Then calculate bootstrap p-value $p* = \frac{1}{n}\sum_{b=1}^{B} 1\{supD^*_b > supD\}$.[5]

2.6 Monte Carlo

We first consider the following data generating process

$$y_{i,n} = \alpha y^m_{i,n}(w) + 3x_{i,n} + \delta_{\alpha_1} y^m_{i,n}(w)1\{z_{i,n} \leq 0\} + \delta_\beta x_{i,n}1\{z_{i,n} \leq 0\} + e_{i,n}, \quad (2.19)$$

where $x_{i,n}$ is a scalar regressor, $z_{i,n}$ is the threshold variable and $e_{i,n}$ is an i.i.d. $N(0,1)$ error.

We set the persistence of the spatial autoregressive coefficient at $\alpha = 0.4$ and explore various experiments that allow to vary the threshold effects $\delta_{\alpha_1} = 0, 0, 1, 0.3, 0.5$ and $\delta_\beta = 0, 1, 2, 3$.[6] We use three alternative predetermined weight matrices: The Toledo spatial matrix WO (98×98) based on the 5 nearest neighbors of 98 census tracts in Toledo, Ohio.For larger sample sizes of $n = 196$ and 392 we use block diagonal matrices with the Toledo spatial matrix as their diagonal blocks.

Tables 2.1-2.5 present the 5th, 50th and 95th quantiles of the distributions of our estimators for three sample size that correspond to the aforementioned weight matrices. Columns 2-4 present the two-stage least squares estimator using the matrix of instrumental variables $Q_n = (X_n, W_nX_n, W_n^2X_n, W_n^3X_n, W_n^4X_n)'$. Columns 5-7 present the first-step GMM estimator using $(Q_n'Q_n)^{-1}$ as weighting matrix and quadratic moments. Columns 8-10 present the second-step GMM estimator using residuals and estimators of the first-step as initial estimates and reiterating the first-step. Columns 11-13 present a third-step estimator using residuals and estimators of the second-step as initial estimates and reiterating the second-step. Each table presents four different panels for different values of the threshold effect δ_β and each panel shows results for four different values of δ_{α_1}. When $\delta_\beta = \delta_{\alpha_1} = 0$ the DGP is the linear model. Finally, we note that our results can be viewed as conservative since we always estimate an unrestricted model regardless of the true DGP.

Table 2.1 shows the results for the estimator of the threshold parameter. Overall, we

[5] $1\{supD^* > supD\}$ takes the value 1 when $supD^* > supD$ and both $supD^*_b$ and $supD$ are positive and 0 otherwise.

[6] The choice of $\alpha = 0.4$ and δ_{α_1} values was made in a way that ensures that Assumption 1 holds.

can see that as the sample size and the threshold effects increase, the 50th quantile approaches the true threshold parameter and the width of the distribution shrinks. While there are efficiency gains as δ_β increases for fixed values of δ_{α_1}, there are no gains in efficiency when δ_{α_1} increases for fixed values of δ_β. We should note that all the estimators appear to center at the true threshold parameter. However, examining the width of the distribution, we see that the GMM does not exhibit efficiency gains over the 2SLS, perhaps due to the slow convergence rate.

Tables 2.2 and 2.3 show the results for the threshold effects of the SAR threshold coefficient δ_{α_1} and the SAR coefficient α, respectively. When $\delta_{\alpha_1} = 0$ the threshold effect of the SAR threshold coefficient δ_{α_1} is estimated accurately regardless of δ_β. However, when $\delta_{\alpha_1} \neq 0$, the estimators of δ_{α_1} shows a finite sample bias that decreases as δ_β increases. Similarly, in the case of the SAR coefficient α, Table 2.3 shows that all estimators accurately estimate α and the 2nd step estimator appears to offer great improvements in efficiency over the first-step estimator in all cases. Importantly, the 2nd step estimator appears to offer great improvements in efficiency for both δ_{α_1} and α over the first-step estimator in all cases. The third-step estimator also provides a small improvement in terms of efficiency but generally not in terms of finite sample bias.

Table 2.4 and 2.5 provide a more nuanced characterization of the performance of the estimators of δ_β and β. In particular, while the 50th quantile approaches the true value and the width of the distribution shrinks when $\delta_\beta = 0$ for all values of δ_{α_1}, there is a finite sample bias for $\delta_\beta \neq 0$. Interestingly, the first-step GMM or 2SLS estimator appears to be better in terms of bias than the second-step estimator and at least as good as the third-step one. As expected, however, the second-step GMM estimator provides substantial efficiency gains. The efficiency gains of the third-step estimator generally appear small. The results for β also show bias for $\delta_\beta \neq 0$ but it generally smaller compared to the bias of the estimators for δ_β. Both larger sample size and third-step estimators appear to reduce the bias to negligible levels.

An alternative possibility is to allow the social network to be different across regimes. Therefore, we consider a second data generating process

$$y_{i,n} = \begin{cases} \alpha_1 \sum_{j\neq i} w_{ij,n} 1\{z_{j,n} \leq 0\} y_{j,n} + \beta_1 x_{i,n} + e_{i,n}, & z_{i,n} \leq 0 \\ \alpha_2 \sum_{j\neq i} w_{ij,n} 1\{z_{j,n} > 0\} y_{j,n} + \beta_2 x_{i,n} + e_{i,n}, & z_{i,n} > 0 \end{cases} \tag{2.20}$$

where $x_{i,n}$ is a scalar regressor, $z_{i,n}$ is the threshold variable and $e_{i,n}$ is an i.i.d. $N(0,1)$ error.

We set the persistence of the spatial autoregressive coefficient $\alpha_2 = 0.4$ and the slope coefficient $\beta_2 = 3$ of the upper regime and vary the spatial autoregressive coefficient α_1 and the slope coefficient β_1 of the lower regime by varying $\delta_{\alpha_1} = \alpha_1 - \alpha_2 = 0, 0, 1, 0.3, 0.5$ and $\delta_\beta = \beta_1 - \beta_2 = 0, 1, 2, 3$. We use two alternative predetermined weight matrices: The Toledo spatial matrix WO (98×98) based on the 5 nearest neighbors of 98 census tracts in Toledo, Ohio.For the larger sample size of $n = 196$ we use a block diagonal matrice with the Toledo spatial matrix as its diagonal blocks.

Tables 2.6-2.10 present the 5th, 50th and 95th quantiles of the distributions of our estimators for two sample size that correspond to the aforementioned weight matrices for Model 2. The same structure as in tables 2.1-2.5 is applied.

Table 2.6 shows the results for the estimator of the threshold parameter. Overall, we can see that as the sample size and the threshold effects increase, the 50th quantile approaches the true threshold parameter. We found similar results in terms of efficiency, as in the first DGP. There are efficiency gains as δ_β increases for fixed values of δ_{α_1}, while there are no gains in efficiency when δ_{α_1} increases for fixed values of δ_β. Although, all the estimators appear to center at the true threshold parameter, we see that the GMM does not exhibit efficiency gains over the 2SLS, perhaps due to the slow convergence rate.

Tables 2.7 and 2.8 show the results for the threshold effects of the SAR threshold coefficient δ_{α_1} and the SAR coefficient α_2, respectively. We see that third step GMM estimator of δ_{α_1} performs better in terms of finite sample bias and provides efficiency gains. Regarding α_2, we can see that the second and third step GMM estimator provide as well gains in terms of finite sample bias and efficiency.

Table 2.9 and 2.10 displays the estimators of δ_β and β. While δ_β is accurately estimated when $\delta_\beta = 0$, there is a bias in the opposite case, which however decreases as δ_β increases. The second and third step GMM estimators provide considerable efficiency gains in comparison with 2sls and first step GMM estimator. In general, the same picture applies for β, although the bias is smaller.

In sum, the results show that our estimators generally accurately estimate the parameters of the model. While the quasi-optimal second-step (or third) estimator does not appear to provide efficiency gains for the estimation of the threshold parameter, it can provide substantial improvements in the efficiency of the estimators for the slope parameters of the model.

2.7 Future Work

In future work we plan to illustrate the empirical relevance of our model by providing an empirical application on intergenerational mobility and social influences. In particular, we will study two possible mechanisms that give rise to poverty traps. First, models of credit constraints such as Galor and Zeira (1993) and Han and Mulligan (2001) suggest that intergenerational mobility threshold-type regressions emerge as distinct intergenerational transmission relationships for constrained and unconstrained families. Second, models of neighborhood effects such as Benabou (1996) and Durlauf (1996a,b) show that parental income plays a role in the quality of the neighborhood in which a child grows up. The quality of neighborhood, in turn, affects future adult income in order to produce threshold-like relationships between parent and offspring income. The idea is that both lagged and contemporaneous feedbacks from the behavior of the members of a neighborhood to the offspring's future outcomes. While lagged feedbacks capture neighborhood effects during childhood, contemporaneous feedbacks capture the idea that neighborhoods constitute an intergenerational transmission mechanism because they provide access to information about employment opportunities. Disadvantaged localities act as barriers to the job opportunities for poor individuals due to lack of hiring networks and general access to information (e.g., Conley and Topa (2001), Topa (2001)). The basic assumption of the two aforementioned papers is that residents of one tract exchange job information with residents of neighboring tracts and that physical distance is an essential determinant of the creation of these networks; the costs of creating and preserving social ties increase with social distance, whereas local institutions such as churches or local businesses help in creating social ties. Therefore, we will treat the census tract as a representative unit of location and assume that a sufficient statistic for the effects of others on a given adult's income is given by a weighted average of income of others, where the weights depend on whether the individual resides in the same census tract or nearby tracts. [7] Since the average commuting distance is equal to 16 miles, we will consider as adjacent tracts the tracts that are in a radius of 16 miles of the centre of the tract. We consider as the tract of residence the tract that individual lived the three years prior the year we have first observed the income used for the construction of permanent income.

The data will be drawn from the Panel Study of Income Dynamics (PSID). PSID is a longitudinal household survey starting in 1968 with a nationally representative sample of over 18,000 individuals living in 5,000 families in the United States. We will use the Survey Research Center national sample and employ measures of the parent's and child's family income, which include the taxable income of all earners in the family, from all sources, and transfer payments. Furthermore, in order to account for the fact

[7] The weights are functions of the distance between the centers of the tracts.

that PSID data are not geographically stratified, we will use Tract contextual data from Neighborhood Change Database (NCDB) such as population density, proportion of black minorities, as well as the proportion of persons over 25 years old who have a bachelors or graduate/professional degree, persons over 25 years old who have completed 9-12 years of schooling and persons over 25 years old who who have completed high school but no college.[8]

Following the literature of threshold regressions, our model can be extended to multiple regimes. Typically, this estimation is performed sequentially, accounting for the order of testing as in Gonzalo and Pitarakis (2002) who suggest repartitioning.

Our model could also be extended to the case where the weight matrices are given exogenously for each regime and hence different socioeconomic matrices are considered across the two regimes. For example, let us assume we study poverty traps and for the formation of the social network, we take into account neighboring tracts. The nature of interactions is intrinsically different for poor and rich people. Therefore, $k-nearest$ neighbors could be considered for the poor, and $m-nearest$ for rich, where $k \neq m$.

2.8 Conclusion

In this chapter we propose a general threshold spatial autoregression model that nests several models including the spatial autoregression model and spatial autoregression model - mixed regression. Using a framework that allows for both fixed and diminishing threshold effects we develop a two-step GMM estimation method that exploit both linear and quadratic moment conditions and study the limiting properties of the estimators of the threshold parameter and slope parameters of spatial lags and regression coefficients. In particular, the first-step estimator is based on an initial matrix of instruments by exploiting the powers of the exogenous spatial weight matrix. and second-step estimator uses regime specific instruments. Finally, we assess the performance of the proposed estimation method using a Monte Carlo simulation.

[8]We would like thank Giulio Zanella who kindly provided us with the NCDB data.

2.9 Tables

Table 2.1: **Model 1 - Threshold Parameter** ($\lambda_0 = 0$)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel A: $\delta_\beta = 0$						
						$\delta_\alpha = 0$						
98	-0.73	-0.03	0.67	-0.71	-0.03	0.63	-0.75	-0.09	0.70	-0.75	0.00	0.69
196	-0.69	0.08	0.66	-0.67	-0.02	0.63	-0.70	0.01	0.71	-0.70	-0.01	0.70
392	-0.65	0.00	0.64	-0.65	-0.05	0.61	-0.70	-0.04	0.69	-0.69	0.00	0.69
						$\delta_\alpha = 0.1$						
98	-0.71	0.02	0.65	-0.71	-0.07	0.61	-0.76	-0.02	0.69	-0.77	-0.06	0.70
196	-0.65	0.01	0.64	-0.65	0.00	0.64	-0.73	0.02	0.70	-0.69	0.05	0.70
392	-0.65	0.00	0.65	-0.64	-0.03	0.61	-0.71	-0.02	0.68	-0.69	-0.03	0.68
						$\delta_\alpha = 0.3$						
98	-0.70	-0.04	0.65	-0.70	-0.05	0.63	-0.77	-0.01	0.68	-0.75	-0.03	0.67
196	-0.67	-0.03	0.66	-0.66	-0.02	0.62	-0.71	0.03	0.68	-0.70	-0.01	0.65
392	-0.66	0.01	0.67	-0.64	-0.05	0.61	-0.69	0.00	0.67	-0.69	-0.03	0.67
						$\delta_\alpha = 0.5$						
98	-0.70	-0.01	0.66	-0.72	-0.06	0.61	-0.70	-0.03	0.65	-0.70	-0.03	0.64
196	-0.67	-0.03	0.65	-0.66	-0.07	0.63	-0.67	0.00	0.67	-0.68	-0.01	0.67
392	-0.64	0.02	0.64	-0.63	0.02	0.62	-0.67	-0.01	0.66	-0.66	-0.02	0.66
						Panel B: $\delta_\beta = 1$						
						$\delta_\alpha = 0$						
98	-0.69	-0.01	0.65	-0.71	-0.02	0.62	-0.72	-0.03	0.64	-0.71	-0.04	0.63
196	-0.64	-0.01	0.60	-0.65	-0.02	0.63	-0.67	0.02	0.65	-0.69	-0.01	0.67
392	-0.63	-0.01	0.62	-0.62	-0.01	0.60	-0.66	-0.03	0.64	-0.68	-0.02	0.62
						$\delta_\alpha = 0.1$						
98	-0.67	-0.02	0.60	-0.69	-0.02	0.60	-0.71	-0.02	0.65	-0.72	-0.02	0.65
196	-0.65	0.00	0.63	-0.67	-0.01	0.64	-0.65	0.02	0.67	-0.68	0.03	0.65
392	-0.60	-0.02	0.60	-0.63	-0.05	0.61	-0.64	-0.03	0.66	-0.65	-0.05	0.63
						$\delta_\alpha = 0.3$						
98	-0.68	-0.01	0.60	-0.71	-0.06	0.60	-0.68	0.00	0.63	-0.70	-0.06	0.64
196	-0.63	0.00	0.62	-0.64	-0.03	0.62	-0.64	0.00	0.63	-0.63	-0.02	0.64
392	-0.62	0.00	0.62	-0.62	-0.03	0.62	-0.66	0.00	0.63	-0.66	0.00	0.63
						$\delta_\alpha = 0.5$						
98	-0.66	-0.02	0.61	-0.66	-0.02	0.61	-0.69	-0.02	0.64	-0.68	-0.01	0.60
196	-0.61	-0.01	0.59	-0.63	-0.06	0.62	-0.63	0.00	0.62	-0.61	0.00	0.65
392	-0.61	-0.01	0.58	-0.61	0.00	0.62	-0.62	-0.01	0.64	-0.65	-0.02	0.62

Model 1 - Threshold Parameter ($\lambda_0 = 0$) (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	-0.64	-0.02	0.53	-0.68	-0.03	0.58	-0.69	-0.03	0.59	-0.71	-0.04	0.58
196	-0.54	0.00	0.54	-0.61	-0.02	0.60	-0.61	-0.01	0.62	-0.60	-0.01	0.59
392	-0.52	-0.01	0.52	-0.59	-0.01	0.59	-0.62	-0.01	0.59	-0.61	-0.01	0.59
						$\delta_\alpha = 0.1$						
98	-0.59	-0.02	0.53	-0.67	-0.02	0.62	-0.65	-0.01	0.59	-0.61	-0.01	0.59
196	-0.53	-0.01	0.52	-0.62	-0.02	0.59	-0.58	0.01	0.62	-0.61	-0.01	0.61
392	-0.47	0.00	0.53	-0.62	-0.04	0.57	-0.62	-0.03	0.60	-0.60	-0.01	0.59
						$\delta_\alpha = 0.3$						
98	-0.58	-0.01	0.56	-0.64	0.00	0.62	-0.62	-0.01	0.59	-0.63	-0.01	0.59
196	-0.56	-0.01	0.51	-0.62	-0.03	0.59	-0.59	-0.01	0.58	-0.60	-0.01	0.58
392	-0.48	-0.01	0.51	-0.61	-0.02	0.58	-0.61	-0.02	0.57	-0.61	-0.01	0.58
						$\delta_\alpha = 0.5$						
98	-0.58	-0.03	0.53	-0.66	-0.03	0.61	-0.61	-0.02	0.57	-0.59	-0.02	0.57
196	-0.57	-0.01	0.57	-0.61	-0.05	0.57	-0.60	-0.02	0.56	-0.59	-0.01	0.57
392	-0.46	0.00	0.50	-0.59	-0.01	0.59	-0.60	-0.01	0.58	-0.58	-0.01	0.56
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	-0.50	-0.01	0.44	-0.63	-0.02	0.59	-0.57	-0.02	0.49	-0.56	-0.02	0.54
196	-0.39	-0.01	0.48	-0.60	-0.02	0.57	-0.57	0.00	0.55	-0.55	0.00	0.55
392	-0.40	0.00	0.32	-0.57	0.00	0.53	-0.56	-0.02	0.56	-0.57	-0.01	0.54
						$\delta_\alpha = 0.1$						
98	-0.48	-0.01	0.46	-0.67	-0.02	0.56	-0.59	-0.01	0.55	-0.59	-0.02	0.54
196	-0.43	-0.01	0.45	-0.62	-0.02	0.59	-0.58	0.00	0.52	-0.56	0.00	0.53
392	-0.36	0.00	0.40	-0.59	-0.02	0.57	-0.58	-0.02	0.54	-0.58	-0.01	0.49
						$\delta_\alpha = 0.3$						
98	-0.51	-0.02	0.49	-0.65	-0.03	0.58	-0.56	-0.01	0.55	-0.58	-0.01	0.56
196	-0.36	-0.01	0.44	-0.58	-0.01	0.56	-0.57	0.00	0.52	-0.53	-0.02	0.52
392	-0.40	0.00	0.31	-0.59	-0.02	0.55	-0.58	0.00	0.59	-0.54	-0.01	0.51
						$\delta_\alpha = 0.5$						
98	-0.48	-0.01	0.44	-0.65	-0.02	0.55	-0.54	-0.01	0.54	-0.59	-0.02	0.52
196	-0.38	-0.01	0.38	-0.60	-0.02	0.57	-0.54	-0.01	0.52	-0.54	-0.01	0.56
392	-0.35	0.00	0.34	-0.56	-0.02	0.56	-0.53	-0.01	0.53	-0.55	-0.01	0.51

Table 2.2: **Model 1 - Threshold Effect of the SAR Coefficient δ_{α_1}**

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
Panel A: $\delta_\beta = 0$												
$\delta_\alpha = 0$												
98	-3.51	-0.06	4.42	-2.35	0.00	2.34	-0.79	-0.01	0.74	-0.57	0.01	0.58
196	-3.95	-0.05	3.55	-2.55	-0.05	2.60	-0.49	0.00	0.51	-0.33	0.00	0.35
392	-3.90	0.04	3.54	-2.62	-0.01	2.82	-0.33	0.00	0.35	-0.24	0.00	0.25
$\delta_\alpha = 0.1$												
98	-3.18	0.07	3.44	-2.44	0.07	2.57	-0.61	0.11	0.76	-0.43	0.13	0.67
196	-3.61	0.15	3.72	-2.50	0.10	2.65	-0.32	0.13	0.64	-0.21	0.12	0.52
392	-3.46	0.32	4.29	-2.58	0.22	2.50	-0.18	0.13	0.49	-0.09	0.12	0.40
$\delta_\alpha = 0.3$												
98	-3.09	0.41	3.57	-2.01	0.39	2.88	-0.29	0.39	1.19	-0.09	0.35	1.00
196	-3.74	0.50	4.06	-2.32	0.39	2.84	-0.09	0.38	1.00	0.07	0.35	0.82
392	-3.93	0.53	3.91	-2.21	0.41	2.72	0.10	0.38	0.86	0.14	0.34	0.74
$\delta_\alpha = 0.5$												
98	-2.74	0.64	3.49	-1.83	0.58	2.82	0.05	0.59	1.47	0.13	0.57	1.33
196	-3.30	0.69	3.93	-1.98	0.75	3.10	0.24	0.62	1.35	0.26	0.57	1.22
392	-2.62	0.87	4.49	-1.80	0.68	3.11	0.28	0.61	1.18	0.31	0.54	1.07
Panel B: $\delta_\beta = 1$												
$\delta_\alpha = 0$												
98	-3.21	0.04	3.10	-2.12	0.04	2.46	-0.57	-0.01	0.63	-0.48	0.00	0.42
196	-3.16	-0.02	3.45	-2.26	-0.02	2.36	-0.38	0.00	0.34	-0.30	0.00	0.26
392	-3.40	0.07	3.08	-2.36	0.05	2.28	-0.23	0.00	0.25	-0.18	0.00	0.19
$\delta_\alpha = 0.1$												
98	-2.99	0.19	3.12	-1.82	0.19	2.31	-0.43	0.12	0.67	-0.29	0.09	0.54
196	-3.31	0.16	3.43	-2.27	0.12	2.38	-0.23	0.12	0.53	-0.15	0.10	0.42
392	-3.00	0.27	3.44	-2.13	0.21	2.47	-0.12	0.11	0.46	-0.06	0.10	0.34
$\delta_\alpha = 0.3$												
98	-2.50	0.43	2.94	-1.84	0.34	2.38	-0.11	0.35	0.99	-0.03	0.33	0.80
196	-2.69	0.28	3.32	-1.88	0.34	2.49	0.05	0.36	0.86	0.10	0.32	0.74
392	-3.05	0.49	3.46	-2.00	0.39	2.44	0.12	0.35	0.73	0.15	0.32	0.68
$\delta_\alpha = 0.5$												
98	-2.29	0.56	3.07	-1.62	0.54	2.68	0.12	0.57	1.26	0.20	0.53	1.19
196	-2.58	0.62	3.44	-1.64	0.55	2.40	0.29	0.58	1.18	0.30	0.54	1.09
392	-2.51	0.77	3.40	-1.75	0.65	2.63	0.33	0.57	1.05	0.31	0.53	1.02

Model 1 - Threshold Effect of the SAR Coefficient δ_{α_1} (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	-2.71	0.04	2.73	-1.83	0.04	2.27	-0.53	-0.01	0.45	-0.43	-0.01	0.36
196	-2.71	-0.03	2.65	-2.18	-0.04	1.74	-0.32	0.00	0.30	-0.21	0.00	0.22
392	-2.91	0.04	2.58	-2.19	0.02	2.05	-0.23	-0.01	0.21	-0.16	0.00	0.17
						$\delta_\alpha = 0.1$						
98	-2.30	0.12	2.41	-1.76	0.07	2.16	-0.31	0.11	0.62	-0.25	0.10	0.46
196	-2.99	0.13	2.78	-2.01	0.12	2.09	-0.15	0.11	0.48	-0.10	0.10	0.35
392	-2.50	0.18	2.92	-1.91	0.21	2.28	-0.07	0.10	0.37	-0.05	0.11	0.29
						$\delta_\alpha = 0.3$						
98	-2.62	0.28	2.77	-1.70	0.30	2.01	-0.07	0.32	0.84	0.02	0.31	0.74
196	-2.12	0.30	2.82	-1.61	0.35	2.17	0.07	0.33	0.71	0.11	0.32	0.64
392	-2.58	0.39	2.79	-1.80	0.38	2.27	0.15	0.33	0.67	0.17	0.32	0.61
						$\delta_\alpha = 0.5$						
98	-2.03	0.50	2.93	-1.59	0.46	2.34	0.15	0.53	1.26	0.24	0.53	1.06
196	-2.06	0.57	2.88	-1.45	0.55	2.44	0.29	0.55	1.02	0.30	0.53	0.96
392	-2.27	0.63	3.19	-1.60	0.62	2.46	0.32	0.54	0.98	0.34	0.51	0.92
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	-2.44	-0.02	2.15	-1.85	-0.02	1.79	-0.39	-0.01	0.37	-0.27	0.00	0.31
196	-2.41	0.00	2.80	-1.82	-0.02	1.99	-0.26	0.00	0.26	-0.20	0.00	0.21
392	-2.48	0.04	2.38	-1.91	0.04	1.88	-0.20	0.00	0.19	-0.15	0.00	0.15
						$\delta_\alpha = 0.1$						
98	-2.48	0.09	2.20	-2.07	0.12	1.98	-0.28	0.12	0.56	-0.19	0.11	0.46
196	-2.74	0.10	2.43	-1.92	0.12	1.94	-0.16	0.11	0.44	-0.09	0.10	0.30
392	-2.02	0.14	2.55	-1.66	0.18	2.23	-0.08	0.10	0.34	-0.04	0.10	0.28
						$\delta_\alpha = 0.3$						
98	-2.02	0.33	2.55	-1.38	0.30	2.20	-0.02	0.32	0.81	0.04	0.32	0.72
196	-2.02	0.29	2.35	-1.47	0.34	2.24	0.11	0.32	0.68	0.13	0.31	0.64
392	-2.18	0.35	2.46	-1.78	0.34	2.12	0.16	0.32	0.64	0.17	0.31	0.57
						$\delta_\alpha = 0.5$						
98	-1.64	0.46	2.40	-1.47	0.43	2.04	0.20	0.52	0.99	0.24	0.51	1.03
196	-1.66	0.49	2.43	-1.26	0.49	2.22	0.29	0.53	1.01	0.33	0.51	0.95
392	-1.94	0.56	2.60	-1.44	0.50	2.33	0.35	0.54	0.93	0.37	0.52	0.88

Table 2.3: **Model 1 - SAR Coefficient** $\alpha = 0.4$

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel A: $\delta_\beta = 0$						
						$\delta_\alpha = 0$						
98	-1.64	0.41	2.18	-0.75	0.39	1.67	-0.02	0.39	0.79	0.09	0.39	0.70
196	-1.61	0.43	2.57	-0.90	0.43	1.72	0.11	0.39	0.68	0.19	0.39	0.60
392	-1.60	0.38	2.30	-0.99	0.39	1.76	0.22	0.40	0.59	0.25	0.40	0.53
						$\delta_\alpha = 0.1$						
98	-1.34	0.40	2.23	-0.92	0.40	1.67	-0.01	0.39	0.72	0.04	0.40	0.68
196	-1.51	0.39	2.44	-0.81	0.42	1.80	0.05	0.39	0.63	0.13	0.39	0.57
392	-1.71	0.32	2.25	-0.94	0.35	1.74	0.15	0.39	0.54	0.23	0.40	0.50
						$\delta_\alpha = 0.3$						
98	-1.34	0.39	2.16	-0.87	0.38	1.56	-0.16	0.36	0.68	-0.07	0.38	0.63
196	-1.57	0.32	2.31	-1.03	0.36	1.78	-0.08	0.36	0.59	0.04	0.38	0.54
392	-1.60	0.30	2.18	-0.95	0.37	1.53	0.01	0.37	0.54	0.09	0.39	0.51
						$\delta_\alpha = 0.5$						
98	-1.22	0.38	2.00	-0.81	0.40	1.54	-0.21	0.37	0.68	-0.15	0.38	0.63
196	-1.57	0.31	2.25	-1.00	0.33	1.60	-0.27	0.35	0.60	-0.08	0.37	0.57
392	-1.68	0.24	1.92	-0.97	0.33	1.56	-0.09	0.37	0.55	-0.02	0.39	0.54
						Panel B: $\delta_\beta = 1$						
						$\delta_\alpha = 0$						
98	-1.21	0.39	2.13	-0.81	0.37	1.49	0.07	0.39	0.74	0.14	0.39	0.64
196	-1.35	0.41	2.08	-0.79	0.41	1.73	0.19	0.39	0.61	0.24	0.40	0.55
392	-1.29	0.37	2.09	-0.88	0.38	1.61	0.24	0.40	0.54	0.28	0.40	0.51
						$\delta_\alpha = 0.1$						
98	-1.25	0.38	1.95	-0.80	0.36	1.35	0.04	0.39	0.68	0.14	0.40	0.62
196	-1.42	0.38	2.14	-0.82	0.41	1.64	0.13	0.39	0.58	0.19	0.40	0.55
392	-1.35	0.32	1.89	-0.91	0.35	1.54	0.20	0.39	0.52	0.28	0.40	0.49
						$\delta_\alpha = 0.3$						
98	-0.95	0.34	1.83	-0.72	0.39	1.43	-0.07	0.37	0.63	0.10	0.39	0.59
196	-1.11	0.42	1.94	-0.82	0.39	1.46	0.03	0.38	0.55	0.12	0.39	0.53
392	-1.37	0.34	2.16	-0.77	0.37	1.53	0.10	0.39	0.51	0.14	0.39	0.50
						$\delta_\alpha = 0.5$						
98	-1.09	0.40	1.79	-0.70	0.39	1.51	-0.11	0.37	0.64	-0.04	0.38	0.59
196	-1.19	0.36	2.03	-0.67	0.43	1.55	-0.15	0.37	0.56	0.04	0.38	0.54
392	-1.16	0.30	1.84	-0.82	0.35	1.48	-0.04	0.39	0.54	0.04	0.39	0.54

Model 1 - SAR Coefficient $\alpha = 0.4$ (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	-0.82	0.39	1.87	-0.78	0.38	1.43	0.14	0.40	0.71	0.18	0.41	0.64
196	-0.94	0.41	1.84	-0.48	0.42	1.55	0.21	0.40	0.59	0.25	0.39	0.52
392	-0.91	0.38	1.82	-0.62	0.38	1.51	0.26	0.40	0.53	0.29	0.39	0.50
						$\delta_\alpha = 0.1$						
98	-0.78	0.40	1.62	-0.62	0.40	1.34	0.10	0.38	0.61	0.17	0.39	0.58
196	-0.88	0.39	1.85	-0.63	0.40	1.50	0.15	0.39	0.56	0.25	0.40	0.53
392	-1.18	0.36	1.74	-0.66	0.35	1.41	0.24	0.40	0.51	0.28	0.40	0.48
						$\delta_\alpha = 0.3$						
98	-0.94	0.42	1.81	-0.53	0.41	1.39	0.05	0.38	0.60	0.13	0.39	0.57
196	-0.85	0.41	1.62	-0.53	0.40	1.35	0.12	0.39	0.55	0.15	0.39	0.52
392	-1.03	0.36	1.77	-0.61	0.38	1.47	0.16	0.39	0.51	0.17	0.39	0.50
						$\delta_\alpha = 0.5$						
98	-0.93	0.42	1.72	-0.53	0.44	1.47	-0.07	0.39	0.60	0.05	0.39	0.57
196	-0.80	0.37	1.63	-0.55	0.40	1.36	0.06	0.38	0.54	0.06	0.39	0.53
392	-0.96	0.35	1.83	-0.75	0.36	1.49	-0.02	0.39	0.53	0.11	0.40	0.52
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	-0.78	0.41	1.61	-0.51	0.42	1.34	0.17	0.40	0.61	0.20	0.40	0.57
196	-0.92	0.40	1.69	-0.55	0.42	1.35	0.23	0.40	0.56	0.25	0.40	0.51
392	-0.76	0.38	1.65	-0.51	0.38	1.44	0.27	0.40	0.52	0.30	0.40	0.49
						$\delta_\alpha = 0.1$						
98	-0.70	0.41	1.64	-0.58	0.39	1.49	0.12	0.39	0.60	0.20	0.39	0.57
196	-0.76	0.41	1.89	-0.47	0.40	1.39	0.17	0.40	0.55	0.26	0.40	0.51
392	-0.81	0.37	1.46	-0.80	0.36	1.34	0.26	0.40	0.51	0.31	0.40	0.48
						$\delta_\alpha = 0.3$						
98	-0.79	0.38	1.49	-0.48	0.40	1.30	0.07	0.39	0.58	0.11	0.38	0.55
196	-0.62	0.41	1.52	-0.57	0.39	1.31	0.11	0.39	0.52	0.20	0.39	0.50
392	-0.66	0.38	1.62	-0.45	0.40	1.54	0.17	0.40	0.50	0.22	0.39	0.49
						$\delta_\alpha = 0.5$						
98	-0.59	0.43	1.40	-0.44	0.44	1.37	0.06	0.39	0.59	0.11	0.39	0.56
196	-0.62	0.42	1.53	-0.44	0.42	1.28	0.07	0.39	0.54	0.09	0.40	0.52
392	-0.70	0.38	1.52	-0.49	0.41	1.40	0.12	0.39	0.51	0.18	0.40	0.50

Table 2.4: **Model 1 - Slope Coefficient δ_β**

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel A: $\delta_\beta = 0$						
						$\delta_\alpha = 0$						
98	-1.81	0.03	1.93	-3.10	0.04	3.32	-0.72	0.02	0.69	-0.70	0.01	0.76
196	-1.79	-0.01	1.78	-3.48	0.05	3.58	-0.54	0.01	0.49	-0.52	0.01	0.51
392	-1.50	-0.04	1.49	-3.39	0.03	3.29	-0.36	-0.02	0.33	-0.36	-0.01	0.35
						$\delta_\alpha = 0.1$						
98	-2.05	-0.04	1.82	-3.16	0.11	3.32	-0.75	0.00	0.73	-0.75	0.02	0.75
196	-1.73	-0.05	1.62	-3.28	-0.01	3.48	-0.48	0.01	0.51	-0.49	0.00	0.49
392	-1.58	0.01	1.60	-3.50	-0.03	3.30	-0.35	0.00	0.35	-0.35	0.00	0.31
						$\delta_\alpha = 0.3$						
98	-1.90	0.02	2.06	-3.31	0.09	3.34	-0.69	-0.01	0.72	-0.68	0.01	0.68
196	-1.54	0.00	1.69	-3.76	-0.13	3.02	-0.51	0.01	0.45	-0.49	0.00	0.45
392	-1.54	-0.03	1.47	-3.31	0.02	3.25	-0.33	-0.01	0.32	-0.32	-0.01	0.29
						$\delta_\alpha = 0.5$						
98	-1.93	-0.02	2.03	-3.27	-0.01	3.25	-0.73	-0.01	0.72	-0.71	0.00	0.72
196	-1.79	0.00	1.85	-3.28	-0.01	3.31	-0.49	0.01	0.46	-0.43	0.01	0.45
392	-1.46	0.00	1.55	-3.49	-0.05	3.33	-0.32	-0.01	0.33	-0.30	0.00	0.30
						Panel B: $\delta_\beta = 1$						
						$\delta_\alpha = 0$						
98	-0.88	0.96	2.86	-2.50	1.04	4.34	0.54	1.18	2.56	0.46	1.09	2.32
196	-0.76	0.93	2.54	-2.62	1.06	4.23	0.67	1.19	2.22	0.58	1.07	2.20
392	-0.85	0.96	2.48	-2.47	1.24	4.24	0.73	1.19	2.04	0.66	1.07	2.02
						$\delta_\alpha = 0.1$						
98	-0.88	1.05	2.87	-2.25	1.09	4.14	0.52	1.18	2.44	0.48	1.10	2.35
196	-0.78	0.93	2.58	-2.27	1.09	4.49	0.66	1.17	2.19	0.61	1.07	2.08
392	-0.66	0.95	2.41	-2.23	1.15	4.32	0.71	1.20	2.05	0.68	1.08	2.01
						$\delta_\alpha = 0.3$						
98	-1.11	0.96	2.79	-2.40	0.95	4.19	0.53	1.15	2.31	0.48	1.11	2.30
196	-0.85	1.00	2.68	-2.65	1.06	4.43	0.66	1.18	2.24	0.63	1.09	2.10
392	-0.67	0.91	2.42	-2.58	1.13	4.42	0.72	1.16	2.04	0.67	1.07	2.05
						$\delta_\alpha = 0.5$						
98	-1.18	0.98	3.02	-2.26	1.03	4.36	0.49	1.13	2.33	0.46	1.07	2.12
196	-0.81	1.00	2.78	-2.37	1.05	4.08	0.65	1.17	2.16	0.61	1.08	2.02
392	-0.52	0.90	2.44	-2.57	0.99	4.12	0.73	1.14	1.98	0.68	1.08	1.99

Model 1 - Slope Coefficient δ_β (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	0.16	1.88	3.52	-1.74	2.04	5.04	1.36	2.18	4.09	1.33	2.05	4.05
196	0.41	1.96	3.55	-1.65	2.09	5.48	1.53	2.22	3.89	1.47	2.10	3.74
392	0.32	1.92	3.57	-1.70	2.26	5.66	1.61	2.20	3.76	1.49	2.08	3.68
						$\delta_\alpha = 0.1$						
98	-0.25	1.94	3.64	-2.02	2.00	5.40	1.47	2.23	4.06	1.33	2.11	3.69
196	0.12	1.87	3.36	-1.78	2.06	5.34	1.51	2.21	3.87	1.45	2.07	3.63
392	0.56	1.93	3.35	-2.18	2.07	5.26	1.54	2.20	3.62	1.51	2.12	3.71
						$\delta_\alpha = 0.3$						
98	-0.31	1.87	3.80	-1.85	2.02	5.56	1.40	2.19	3.94	1.38	2.12	3.87
196	0.09	1.95	3.63	-2.00	2.01	4.98	1.53	2.16	3.69	1.47	2.08	3.58
392	0.32	1.91	3.37	-1.58	2.24	5.74	1.57	2.19	3.71	1.48	2.08	3.62
						$\delta_\alpha = 0.5$						
98	-0.15	1.94	4.02	-2.21	2.01	5.49	1.39	2.14	3.93	1.38	2.08	3.68
196	-0.01	1.91	3.75	-2.01	2.09	5.13	1.56	2.17	3.69	1.46	2.08	3.58
392	0.42	1.89	3.35	-1.96	2.05	5.38	1.55	2.16	3.55	1.53	2.06	3.49
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	1.09	2.94	4.87	-0.79	3.08	6.47	2.29	3.21	5.42	2.26	3.08	5.51
196	1.18	2.87	4.69	-0.93	3.09	6.68	2.39	3.20	5.23	2.30	3.05	4.91
392	1.12	2.92	4.64	-0.85	3.30	6.95	2.47	3.23	5.19	2.36	3.05	4.92
						$\delta_\alpha = 0.1$						
98	0.97	2.93	4.96	-1.23	3.11	6.57	2.25	3.20	5.55	2.24	3.13	5.14
196	1.09	2.93	4.70	-1.01	3.11	6.74	2.47	3.21	5.24	2.32	3.06	5.05
392	1.29	2.90	4.53	-1.71	3.02	6.58	2.41	3.18	5.22	2.37	3.06	5.10
						$\delta_\alpha = 0.3$						
98	0.61	2.82	4.85	-1.31	2.94	6.40	2.27	3.18	5.24	2.20	3.07	5.47
196	1.17	2.89	4.77	-0.58	3.12	7.05	2.39	3.18	5.24	2.32	3.10	5.17
392	1.33	2.93	4.55	-0.54	3.26	6.94	2.47	3.18	5.56	2.32	3.05	4.98
						$\delta_\alpha = 0.5$						
98	0.62	2.84	4.96	-1.16	3.03	6.92	2.29	3.13	4.96	2.29	3.10	5.46
196	1.21	2.96	4.81	-1.18	3.14	6.76	2.43	3.18	5.19	2.32	3.09	5.26
392	1.30	2.92	4.57	-1.13	3.14	6.78	2.48	3.18	5.19	2.36	3.08	5.04

Table 2.5: **Model 1 - Coefficient β**

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
Panel A: $\delta_\beta = 0$												
$\delta_\alpha = 0$												
98	1.97	2.99	4.00	1.31	2.99	4.69	2.62	3.00	3.37	2.60	3.00	3.42
196	2.03	3.00	3.98	1.37	2.98	4.75	2.70	3.00	3.30	2.71	2.99	3.30
392	2.22	3.01	3.79	1.35	2.97	4.76	2.81	3.00	3.19	2.80	3.00	3.19
$\delta_\alpha = 0.1$												
98	2.03	3.02	4.15	1.42	2.95	4.70	2.57	3.01	3.43	2.57	3.00	3.43
196	2.21	3.02	3.94	1.23	3.00	4.74	2.72	3.00	3.28	2.72	3.00	3.29
392	2.25	3.00	3.81	1.38	3.02	4.85	2.82	3.01	3.20	2.83	3.00	3.20
$\delta_\alpha = 0.3$												
98	1.93	2.98	3.99	1.27	2.91	4.69	2.55	2.99	3.42	2.60	2.99	3.37
196	2.08	2.99	3.80	1.49	3.08	4.83	2.73	2.99	3.31	2.75	2.99	3.26
392	2.27	3.02	3.83	1.37	2.98	4.83	2.81	3.00	3.19	2.84	3.00	3.17
$\delta_\alpha = 0.5$												
98	1.85	3.01	3.98	1.29	3.01	4.64	2.57	3.00	3.41	2.60	3.00	3.38
196	2.15	2.99	3.96	1.26	3.00	4.68	2.73	3.00	3.31	2.74	3.00	3.25
392	2.17	3.01	3.83	1.34	3.03	4.77	2.83	3.01	3.19	2.82	3.01	3.16
Panel B: $\delta_\beta = 1$												
$\delta_\alpha = 0$												
98	1.99	3.06	3.95	1.29	3.03	4.67	1.97	2.96	3.34	2.13	2.98	3.39
196	2.22	3.07	3.83	1.25	3.01	4.78	2.05	2.94	3.25	2.18	2.98	3.29
392	2.18	3.05	3.90	1.22	2.95	4.86	2.16	2.97	3.22	2.30	2.99	3.28
$\delta_\alpha = 0.1$												
98	1.88	3.02	3.98	1.21	3.00	4.79	1.97	2.95	3.32	2.14	2.98	3.37
196	2.11	3.03	3.99	1.12	3.00	4.75	2.04	2.95	3.27	2.26	2.98	3.29
392	2.22	3.04	3.81	1.13	2.98	4.73	2.15	2.97	3.25	2.23	2.99	3.27
$\delta_\alpha = 0.3$												
98	1.95	3.05	4.08	1.29	3.10	4.93	2.04	2.95	3.37	2.21	2.98	3.39
196	2.06	3.02	4.03	0.97	3.03	4.75	2.10	2.95	3.28	2.25	2.97	3.28
392	2.23	3.04	3.90	1.28	3.00	4.71	2.15	2.98	3.23	2.21	2.99	3.26
$\delta_\alpha = 0.5$												
98	1.96	3.04	4.17	1.31	3.03	4.80	2.14	2.97	3.37	2.23	2.98	3.36
196	1.95	3.00	3.87	1.33	3.06	4.63	2.14	2.94	3.27	2.23	2.98	3.28
392	2.23	3.05	3.78	1.28	3.03	4.85	2.22	2.98	3.25	2.26	3.00	3.27

Model 1 - Coefficient β (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	2.05	3.04	4.12	1.29	3.05	4.82	1.55	2.97	3.53	1.56	2.99	3.61
196	2.11	3.04	3.91	1.05	2.96	4.87	1.52	2.95	3.39	1.80	2.98	3.46
392	2.20	3.04	3.86	1.11	2.95	4.73	1.57	2.98	3.36	1.69	2.99	3.45
						$\delta_\alpha = 0.1$						
98	2.02	3.08	4.14	1.04	3.09	5.00	1.53	2.93	3.40	1.70	2.97	3.46
196	2.21	3.07	3.94	1.15	3.04	4.78	1.41	2.95	3.41	1.87	2.98	3.49
392	2.27	3.02	3.75	1.08	3.07	5.08	1.67	2.99	3.43	1.61	2.99	3.47
						$\delta_\alpha = 0.3$						
98	1.92	3.08	4.21	0.92	3.05	4.87	1.52	2.95	3.44	1.67	2.97	3.48
196	2.11	3.03	3.99	1.33	3.02	4.93	1.72	2.97	3.40	1.82	2.98	3.47
392	2.24	3.05	3.85	1.01	2.96	4.77	1.72	2.98	3.43	1.74	2.99	3.47
						$\delta_\alpha = 0.5$						
98	1.97	3.04	4.15	1.03	3.09	5.06	1.68	2.99	3.50	1.75	2.99	3.50
196	2.06	3.06	4.09	1.24	3.06	4.98	1.70	2.98	3.40	1.83	2.99	3.45
392	2.24	3.05	3.83	1.15	3.04	4.94	1.74	2.98	3.42	1.89	3.00	3.42
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	1.93	3.04	4.14	1.04	2.99	4.79	1.54	2.96	3.44	1.49	2.98	3.61
196	2.05	3.06	4.10	1.05	3.01	5.01	1.28	2.96	3.49	1.61	2.98	3.59
392	2.13	3.04	3.94	0.84	2.94	4.94	1.33	2.98	3.50	1.51	3.00	3.61
						$\delta_\alpha = 0.1$						
98	1.87	3.03	4.27	1.01	3.03	5.03	1.04	2.96	3.63	1.57	2.99	3.61
196	1.98	3.04	4.15	0.97	3.04	4.91	1.26	2.96	3.48	1.39	2.99	3.58
392	2.15	3.04	3.93	1.03	3.08	5.36	1.45	2.98	3.58	1.64	3.00	3.58
						$\delta_\alpha = 0.3$						
98	1.92	3.10	4.26	1.16	3.08	5.16	1.34	2.95	3.57	1.32	2.97	3.62
196	2.05	3.05	3.98	0.85	3.01	4.72	1.20	2.96	3.58	1.43	2.98	3.57
392	2.13	3.03	3.93	0.89	2.95	4.74	1.07	2.98	3.52	1.51	3.00	3.58
						$\delta_\alpha = 0.5$						
98	1.95	3.08	4.29	0.99	3.04	5.03	1.52	2.94	3.58	1.31	2.98	3.53
196	1.97	3.05	4.05	0.81	3.02	5.03	1.33	2.97	3.50	1.40	2.99	3.58
392	2.17	3.04	3.96	0.84	2.99	5.01	1.37	2.98	3.47	1.59	3.00	3.61

Table 2.6: **Model 2 - Threshold Parameter** ($\lambda_0 = 0$)

n	5th	50th 2SLS	95th	5th	50th 1st step GMM	95th	5th	50th 2nd step GMM	95th	5th	50th 3rd step GMM	95th
						Panel A: $\delta_\beta = 0$						
						$\delta_\alpha = 0$						
98	-0.66	-0.02	0.62	-0.65	-0.02	0.63	-0.69	-0.01	0.62	-0.66	-0.05	0.63
196	-0.62	0.00	0.61	-0.58	0.01	0.65	-0.61	-0.01	0.62	-0.60	0.00	0.62
						$\delta_\alpha = 0.1$						
98	-0.69	-0.03	0.63	-0.68	-0.04	0.59	-0.65	-0.01	0.60	-0.63	-0.01	0.61
196	-0.63	0.00	0.63	-0.65	0.01	0.61	-0.61	0.00	0.65	-0.63	0.01	0.64
						$\delta_\alpha = 0.3$						
98	-0.65	-0.03	0.60	-0.64	-0.02	0.58	-0.63	-0.02	0.57	-0.63	-0.02	0.58
196	-0.61	0.00	0.61	-0.57	0.02	0.62	-0.61	0.01	0.60	-0.60	0.02	0.64
						$\delta_\alpha = 0.5$						
98	-0.64	-0.02	0.62	-0.62	0.00	0.57	-0.56	-0.01	0.55	-0.54	0.00	0.56
196	-0.59	0.01	0.59	-0.61	0.01	0.62	-0.56	0.01	0.62	-0.58	0.01	0.63
						Panel B: $\delta_\beta = 1$						
						$\delta_\alpha = 0$						
98	-0.66	-0.01	0.57	-0.65	-0.01	0.62	-0.63	-0.04	0.63	-0.64	-0.02	0.59
196	-0.57	0.00	0.54	-0.58	0.00	0.61	-0.59	0.00	0.61	-0.61	-0.01	0.58
						$\delta_\alpha = 0.1$						
98	-0.66	-0.03	0.60	-0.64	-0.03	0.59	-0.62	-0.02	0.55	-0.62	-0.01	0.58
196	-0.59	-0.01	0.56	-0.62	0.00	0.59	-0.63	0.00	0.60	-0.60	0.00	0.61
						$\delta_\alpha = 0.3$						
98	-0.64	-0.02	0.52	-0.61	-0.02	0.58	-0.59	-0.02	0.55	-0.59	-0.02	0.59
196	-0.56	0.00	0.55	-0.58	0.00	0.60	-0.57	0.00	0.60	-0.54	0.00	0.61
						$\delta_\alpha = 0.5$						
98	-0.61	-0.03	0.51	-0.61	-0.01	0.55	-0.53	-0.01	0.53	-0.56	-0.01	0.51
196	-0.55	0.00	0.56	-0.58	0.01	0.58	-0.56	0.00	0.55	-0.56	0.01	0.59

Model 2 - Threshold Parameter ($\lambda_0 = 0$) (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	-0.58	-0.02	0.50	-0.63	-0.03	0.56	-0.62	-0.01	0.55	-0.64	-0.02	0.55
196	-0.48	-0.01	0.45	-0.60	0.00	0.61	-0.61	0.00	0.57	-0.63	-0.01	0.56
						$\delta_\alpha = 0.1$						
98	-0.53	-0.03	0.48	-0.64	-0.02	0.60	-0.61	-0.01	0.57	-0.61	-0.02	0.52
196	-0.48	0.00	0.48	-0.58	0.00	0.58	-0.62	-0.01	0.54	-0.57	0.00	0.57
						$\delta_\alpha = 0.3$						
98	-0.55	-0.03	0.51	-0.61	-0.02	0.57	-0.53	-0.01	0.52	-0.53	-0.01	0.50
196	-0.49	-0.01	0.44	-0.57	-0.01	0.55	-0.55	0.00	0.58	-0.51	0.00	0.61
						$\delta_\alpha = 0.5$						
98	-0.47	-0.02	0.48	-0.53	-0.02	0.50	-0.47	-0.02	0.45	-0.44	-0.01	0.46
196	-0.44	0.00	0.47	-0.54	0.00	0.62	-0.53	0.00	0.55	-0.50	0.00	0.59
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	-0.50	-0.02	0.37	-0.62	-0.02	0.56	-0.59	-0.01	0.51	-0.58	-0.01	0.53
196	-0.39	-0.01	0.33	-0.54	-0.01	0.56	-0.58	0.00	0.56	-0.54	-0.01	0.56
						$\delta_\alpha = 0.1$						
98	-0.45	-0.02	0.42	-0.63	-0.03	0.56	-0.59	-0.01	0.52	-0.55	-0.01	0.56
196	-0.35	-0.01	0.38	-0.57	0.00	0.56	-0.57	0.00	0.54	-0.56	0.00	0.54
						$\delta_\alpha = 0.3$						
98	-0.47	-0.01	0.39	-0.58	-0.01	0.57	-0.55	-0.01	0.48	-0.48	-0.01	0.47
196	-0.38	-0.01	0.31	-0.51	0.00	0.55	-0.52	0.00	0.56	-0.51	0.00	0.55
						$\delta_\alpha = 0.5$						
98	-0.39	-0.01	0.45	-0.56	-0.01	0.51	-0.45	-0.02	0.41	-0.46	-0.02	0.46
196	-0.36	-0.01	0.35	-0.54	0.00	0.56	-0.46	0.00	0.50	-0.50	0.00	0.58

Table 2.7: **Model 2 - Threshold Effect of the SAR Coefficient δ_{α_1}**

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
	2SLS			1st step GMM			2nd step GMM			3rd step GMM		
Panel A: $\delta_\beta = 0$												
$\delta_\alpha = 0$												
98	-2.20	-0.01	2.41	-1.71	0.00	1.90	-0.97	0.01	1.01	-0.63	0.00	0.80
196	-3.67	0.06	3.89	-2.68	-0.03	2.42	-1.50	0.00	1.42	-1.08	-0.02	1.02
$\delta_\alpha = 0.1$												
98	-2.35	0.10	2.51	-1.81	0.11	1.73	-0.79	0.10	1.17	-0.52	0.10	0.99
196	-3.54	0.18	4.22	-2.44	0.17	2.66	-1.30	0.11	1.70	-0.87	0.09	0.96
$\delta_\alpha = 0.3$												
98	-2.21	0.25	2.64	-1.42	0.28	2.00	-0.34	0.34	1.56	-0.31	0.32	1.44
196	-3.43	0.30	3.71	-2.36	0.26	2.82	-0.76	0.31	1.63	-0.54	0.31	1.50
$\delta_\alpha = 0.5$												
98	-1.73	0.49	2.78	-1.21	0.45	2.03	-0.25	0.52	1.73	-0.08	0.49	1.48
196	-3.28	0.43	3.65	-2.27	0.49	2.90	-0.55	0.53	2.66	-0.35	0.49	2.06
Panel B: $\delta_\beta = 1$												
$\delta_\alpha = 0$												
98	-2.08	-0.04	2.21	-1.56	0.00	1.50	-0.64	0.03	0.86	-0.49	0.02	0.67
196	-3.16	0.02	3.28	-2.22	0.07	2.36	-0.96	0.01	0.91	-0.66	0.02	0.63
$\delta_\alpha = 0.1$												
98	-1.89	0.20	2.50	-1.58	0.17	1.81	-0.60	0.12	0.99	-0.48	0.11	0.91
196	-2.94	0.15	3.60	-2.13	0.16	2.36	-0.79	0.14	1.23	-0.58	0.12	0.92
$\delta_\alpha = 0.3$												
98	-1.99	0.27	2.37	-1.44	0.27	1.85	-0.30	0.32	1.50	-0.14	0.31	1.19
196	-2.60	0.34	3.44	-1.73	0.33	2.54	-0.55	0.35	1.54	-0.33	0.33	1.38
$\delta_\alpha = 0.5$												
98	-2.00	0.42	2.64	-1.19	0.46	2.17	-0.08	0.51	1.52	0.03	0.51	1.48
196	-2.85	0.51	3.63	-2.01	0.47	2.75	-0.33	0.55	1.92	-0.08	0.53	2.17

Model 2 - Threshold Effect of the SAR Coefficient δ_{α_1} (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
Panel C: $\delta_\beta = 2$												
$\delta_\alpha = 0$												
98	-1.69	0.00	1.92	-1.52	0.04	1.58	-0.58	0.01	0.80	-0.39	0.01	0.54
196	-2.49	-0.02	2.62	-2.11	0.06	1.95	-0.63	0.02	0.93	-0.53	0.02	0.63
$\delta_\alpha = 0.1$												
98	-1.72	0.07	1.96	-1.36	0.08	1.56	-0.50	0.11	0.95	-0.33	0.11	0.75
196	-2.65	0.07	3.12	-2.18	0.14	2.31	-0.60	0.15	1.08	-0.44	0.13	0.84
$\delta_\alpha = 0.3$												
98	-1.77	0.27	2.30	-1.35	0.28	1.88	-0.20	0.33	1.22	-0.09	0.33	1.34
196	-2.01	0.35	3.10	-1.80	0.38	2.29	-0.42	0.35	1.44	-0.29	0.33	1.12
$\delta_\alpha = 0.5$												
98	-1.79	0.43	2.48	-1.14	0.48	1.90	-0.02	0.52	1.45	0.10	0.52	1.52
196	-2.53	0.57	3.37	-1.63	0.54	2.66	-0.26	0.55	1.88	-0.03	0.53	1.57
Panel D: $\delta_\beta = 3$												
$\delta_\alpha = 0$												
98	-1.76	0.02	1.95	-1.46	-0.03	1.41	-0.46	0.03	0.70	-0.37	0.01	0.56
196	-2.07	0.03	2.48	-2.08	0.05	1.87	-0.53	0.03	0.82	-0.54	0.02	0.64
$\delta_\alpha = 0.1$												
98	-1.57	0.09	1.83	-1.56	0.14	1.76	-0.43	0.12	0.90	-0.35	0.11	0.69
196	-2.51	0.09	2.82	-2.17	0.13	2.16	-0.58	0.13	1.05	-0.49	0.12	0.84
$\delta_\alpha = 0.3$												
98	-1.45	0.31	1.92	-1.21	0.29	1.70	-0.16	0.33	1.17	-0.13	0.31	1.13
196	-1.83	0.32	2.66	-1.79	0.37	2.29	-0.39	0.34	1.27	-0.21	0.33	1.02
$\delta_\alpha = 0.5$												
98	-1.57	0.41	2.44	-1.13	0.44	1.89	-0.06	0.52	1.44	0.18	0.51	1.40
196	-2.26	0.53	3.00	-1.86	0.54	2.47	-0.13	0.53	1.67	-0.02	0.55	1.64

Table 2.8: **Model 2 - SAR Coefficient** $\alpha_2 = 0.4$

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel A: $\delta_\beta = 0$						
						$\delta_\alpha = 0$						
98	-0.77	0.40	1.49	-0.59	0.40	1.28	0.08	0.44	1.10	0.14	0.43	0.93
196	-1.64	0.32	2.29	-0.92	0.34	1.93	-0.34	0.37	1.33	-0.11	0.38	1.29
						$\delta_\alpha = 0.1$						
98	-0.71	0.42	1.59	-0.39	0.40	1.29	0.10	0.44	1.06	0.14	0.42	0.92
196	-1.84	0.35	2.19	-1.02	0.32	1.67	-0.44	0.39	1.15	-0.01	0.39	1.10
						$\delta_\alpha = 0.3$						
98	-0.92	0.44	1.65	-0.58	0.41	1.29	0.00	0.41	0.98	0.10	0.41	0.90
196	-1.51	0.36	2.30	-0.94	0.37	1.82	-0.31	0.36	1.14	-0.21	0.36	0.97
						$\delta_\alpha = 0.5$						
98	-0.86	0.44	1.72	-0.52	0.44	1.38	0.05	0.41	0.96	0.16	0.42	0.93
196	-1.50	0.38	2.25	-1.11	0.37	1.83	-0.76	0.37	1.07	-0.15	0.39	1.02
						Panel B: $\delta_\beta = 1$						
						$\delta_\alpha = 0$						
98	-0.67	0.44	1.69	-0.43	0.38	1.31	0.07	0.42	0.95	0.15	0.41	0.86
196	-1.43	0.34	2.06	-1.01	0.31	1.68	-0.16	0.36	1.12	-0.01	0.37	0.85
						$\delta_\alpha = 0.1$						
98	-1.00	0.38	1.57	-0.52	0.38	1.39	0.07	0.41	0.96	0.15	0.42	0.85
196	-1.86	0.34	2.30	-0.97	0.31	1.82	-0.33	0.37	1.08	-0.04	0.37	0.87
						$\delta_\alpha = 0.3$						
98	-0.81	0.44	1.72	-0.51	0.43	1.51	-0.01	0.41	0.90	0.12	0.42	0.84
196	-1.52	0.35	1.94	-1.07	0.34	1.64	-0.25	0.36	1.05	-0.10	0.36	0.81
						$\delta_\alpha = 0.5$						
98	-0.95	0.44	1.76	-0.61	0.44	1.43	0.00	0.40	0.88	0.15	0.40	0.79
196	-1.75	0.37	2.44	-1.11	0.37	1.81	-0.36	0.38	1.05	-0.27	0.37	0.85

Model 2 - SAR Coefficient $\alpha_2 = 0.4$ (Continued)

n	5th	50th 2SLS	95th	5th	50th 1st step GMM	95th	5th	50th 2nd step GMM	95th	5th	50th 3rd step GMM	95th
Panel C: $\delta_\beta = 2$												
$\delta_\alpha = 0$												
98	-0.87	0.40	1.40	-0.57	0.38	1.31	-0.01	0.42	0.94	0.16	0.41	0.78
196	-1.09	0.36	2.06	-0.91	0.30	1.75	-0.22	0.37	0.90	-0.07	0.37	0.81
$\delta_\alpha = 0.1$												
98	-0.65	0.41	1.56	-0.51	0.41	1.34	0.09	0.42	1.00	0.11	0.40	0.82
196	-1.69	0.40	2.22	-0.97	0.33	1.81	-0.27	0.36	0.94	-0.05	0.38	0.84
$\delta_\alpha = 0.3$												
98	-0.92	0.42	1.68	-0.68	0.41	1.46	-0.02	0.40	0.88	0.09	0.40	0.86
196	-1.34	0.35	1.84	-0.93	0.31	1.75	-0.39	0.36	1.04	-0.13	0.37	0.85
$\delta_\alpha = 0.5$												
98	-0.89	0.45	1.83	-0.57	0.42	1.41	0.00	0.40	0.87	0.10	0.40	0.83
196	-1.52	0.37	2.47	-1.16	0.34	1.76	-0.40	0.38	1.02	-0.09	0.38	0.85
Panel D: $\delta_\beta = 3$												
$\delta_\alpha = 0$												
98	-0.78	0.39	1.60	-0.49	0.41	1.42	0.00	0.40	0.82	0.10	0.41	0.78
196	-1.12	0.35	1.89	-0.96	0.30	1.79	-0.24	0.36	0.86	-0.05	0.38	0.78
$\delta_\alpha = 0.1$												
98	-0.70	0.40	1.53	-0.50	0.39	1.55	-0.07	0.40	0.88	0.09	0.40	0.86
196	-1.46	0.39	2.16	-0.86	0.34	2.02	-0.28	0.38	0.97	-0.10	0.39	0.95
$\delta_\alpha = 0.3$												
98	-0.73	0.38	1.58	-0.65	0.41	1.43	-0.02	0.41	0.93	0.12	0.41	0.85
196	-1.02	0.37	1.89	-1.04	0.30	1.96	-0.30	0.38	1.00	-0.13	0.38	0.82
$\delta_\alpha = 0.5$												
98	-0.99	0.45	1.79	-0.64	0.44	1.55	-0.03	0.40	0.98	0.09	0.40	0.74
196	-1.31	0.38	2.37	-1.11	0.36	2.10	-0.38	0.39	0.98	-0.22	0.38	0.83

Table 2.9: **Model 2 - Slope Coefficient δ_β**

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel A: $\delta_\beta = 0$						
						$\delta_\alpha = 0$						
98	-1.74	0.01	1.68	-2.73	0.03	3.58	-0.77	0.02	0.87	-0.70	0.03	0.78
196	-1.28	-0.01	1.27	-3.01	0.02	2.94	-0.54	0.01	0.50	-0.49	0.01	0.47
						$\delta_\alpha = 0.1$						
98	-1.87	0.00	1.59	-3.21	-0.02	3.02	-0.82	0.01	0.91	-0.75	0.02	0.69
196	-1.41	0.01	1.33	-2.91	-0.02	2.71	-0.56	0.01	0.53	-0.45	0.00	0.48
						$\delta_\alpha = 0.3$						
98	-1.76	0.05	2.01	-3.11	0.01	3.29	-0.78	0.00	0.75	-0.83	0.01	0.72
196	-1.40	0.01	1.40	-3.10	-0.01	3.02	-0.51	-0.01	0.49	-0.48	0.00	0.50
						$\delta_\alpha = 0.5$						
98	-1.79	0.05	1.98	-3.81	-0.06	3.75	-0.82	-0.02	0.80	-0.70	-0.03	0.72
196	-1.18	-0.01	1.45	-3.18	-0.02	3.37	-0.52	0.00	0.56	-0.54	0.00	0.48
						Panel B: $\delta_\beta = 1$						
						$\delta_\alpha = 0$						
98	-0.65	0.95	2.73	-2.27	0.93	4.11	0.46	1.12	2.50	0.39	1.09	2.35
196	-0.34	0.94	2.29	-1.85	0.92	4.24	0.60	1.09	2.00	0.62	1.08	2.09
						$\delta_\alpha = 0.1$						
98	-0.84	0.92	2.51	-2.59	0.96	4.37	0.43	1.08	2.36	0.43	1.07	2.55
196	-0.53	0.92	2.28	-2.07	0.95	4.28	0.60	1.11	2.21	0.63	1.09	2.02
						$\delta_\alpha = 0.3$						
98	-0.86	0.91	2.72	-2.24	0.90	4.33	0.37	1.06	2.33	0.43	1.04	2.27
196	-0.44	0.90	2.31	-2.26	0.93	4.22	0.61	1.09	2.04	0.63	1.10	2.06
						$\delta_\alpha = 0.5$						
98	-1.06	0.95	3.03	-2.92	0.98	4.85	0.37	1.06	2.27	0.30	1.05	2.08
196	-0.46	0.91	2.31	-3.08	0.88	4.33	0.61	1.09	2.16	0.60	1.07	2.08

Model 2 - Slope Coefficient δ_β (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
					Panel C: $\delta_\beta = 2$							
					$\delta_\alpha = 0$							
98	0.15	1.87	3.61	-1.70	1.83	4.96	1.38	2.12	3.94	1.37	2.12	3.88
196	0.62	1.90	3.25	-0.82	1.91	5.54	1.54	2.14	3.91	1.52	2.13	3.91
					$\delta_\alpha = 0.1$							
98	-0.11	1.86	3.55	-2.34	1.81	5.24	1.37	2.14	4.05	1.41	2.13	3.95
196	0.29	1.88	3.33	-1.40	1.83	5.04	1.56	2.13	4.02	1.54	2.13	3.98
					$\delta_\alpha = 0.3$							
98	-0.22	1.80	3.68	-2.79	1.89	5.56	1.38	2.12	3.87	1.36	2.05	3.72
196	0.65	1.91	3.33	-1.13	1.85	5.28	1.58	2.13	3.97	1.60	2.10	3.83
					$\delta_\alpha = 0.5$							
98	-0.47	1.96	4.56	-2.69	1.83	6.16	1.39	2.08	3.57	1.36	2.06	3.58
196	0.36	1.87	3.47	-1.85	1.82	6.00	1.52	2.10	3.84	1.54	2.10	3.84
					Panel D: $\delta_\beta = 3$							
					$\delta_\alpha = 0$							
98	1.00	2.83	4.74	-0.75	2.81	6.25	2.25	3.19	5.50	2.38	3.11	5.55
196	1.34	2.92	4.29	-0.54	2.83	6.38	2.42	3.15	5.68	2.48	3.16	5.61
					$\delta_\alpha = 0.1$							
98	0.94	2.89	4.79	-1.02	2.99	6.78	2.34	3.16	5.73	2.31	3.14	5.88
196	1.14	2.84	4.34	-0.58	2.77	6.80	2.46	3.12	5.92	2.53	3.12	5.58
					$\delta_\alpha = 0.3$							
98	0.64	2.80	4.82	-1.39	2.85	6.37	2.36	3.12	5.82	2.33	3.11	5.48
196	1.41	2.89	4.42	-0.60	2.78	6.89	2.49	3.13	5.60	2.52	3.11	5.40
					$\delta_\alpha = 0.5$							
98	0.24	2.95	5.54	-2.43	2.83	7.78	2.34	3.09	5.25	2.41	3.10	5.17
196	1.13	2.84	4.54	-1.82	2.82	6.63	2.52	3.09	5.45	2.54	3.13	5.52

Table 2.10: **Model 2- Coefficient β**

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
Panel A: $\delta_\beta = 0$												
$\delta_\alpha = 0$												
98	2.11	3.00	3.98	1.19	2.97	4.53	2.55	2.99	3.44	2.59	2.99	3.39
196	2.27	3.00	3.64	1.53	3.00	4.57	2.70	3.00	3.30	2.70	3.00	3.31
$\delta_\alpha = 0.1$												
98	2.16	3.01	3.97	1.51	3.02	4.59	2.55	3.00	3.52	2.55	3.00	3.44
196	2.26	2.98	3.74	1.60	3.03	4.55	2.68	3.00	3.29	2.70	3.00	3.26
$\delta_\alpha = 0.3$												
98	1.89	2.95	3.95	1.28	2.98	4.63	2.53	2.98	3.43	2.55	2.98	3.42
196	2.26	2.99	3.70	1.53	3.01	4.58	2.71	3.01	3.31	2.71	3.01	3.29
$\delta_\alpha = 0.5$												
98	1.92	2.96	4.00	1.19	3.04	4.79	2.57	3.00	3.42	2.58	3.00	3.42
196	2.25	2.99	3.70	1.31	3.02	4.67	2.68	3.00	3.31	2.73	3.00	3.27
Panel B: $\delta_\beta = 1$												
$\delta_\alpha = 0$												
98	2.04	3.05	3.85	1.30	3.07	4.61	1.90	2.97	3.37	2.23	2.97	3.40
196	2.25	3.04	3.73	1.31	3.04	4.48	2.28	2.98	3.30	2.30	2.98	3.29
$\delta_\alpha = 0.1$												
98	2.16	3.05	3.98	1.16	3.04	4.80	2.17	2.98	3.41	2.04	2.98	3.38
196	2.30	3.04	3.80	1.24	3.04	4.59	2.21	2.97	3.27	2.28	2.98	3.27
$\delta_\alpha = 0.3$												
98	2.10	3.03	3.95	1.25	3.04	4.61	2.27	2.98	3.40	2.26	2.98	3.39
196	2.28	3.04	3.72	1.22	3.06	4.66	2.22	2.97	3.29	2.19	2.97	3.27
$\delta_\alpha = 0.5$												
98	1.93	3.02	4.12	0.98	3.03	5.23	2.32	2.99	3.35	2.41	2.99	3.38
196	2.25	3.05	3.78	1.25	3.05	5.07	2.22	2.96	3.26	2.22	2.97	3.26

Model 2 - Coefficient β (Continued)

n	5th	50th	95th	5th	50th	95th	5th	50th	95th	5th	50th	95th
		2SLS			1st step GMM			2nd step GMM			3rd step GMM	
						Panel C: $\delta_\beta = 2$						
						$\delta_\alpha = 0$						
98	2.09	3.08	4.03	1.30	3.14	4.84	1.55	2.97	3.49	1.57	2.98	3.48
196	2.27	3.04	3.82	1.32	3.06	4.39	1.61	2.96	3.35	1.66	2.98	3.40
						$\delta_\alpha = 0.1$						
98	2.09	3.10	4.09	1.34	3.14	4.96	1.72	2.96	3.47	1.70	2.98	3.43
196	2.23	3.07	3.93	1.25	3.10	4.69	1.61	2.98	3.40	1.58	2.97	3.37
						$\delta_\alpha = 0.3$						
98	2.06	3.09	4.21	0.85	3.07	5.43	1.62	2.96	3.41	1.98	2.98	3.44
196	2.24	3.06	3.77	1.09	3.13	4.62	1.41	2.95	3.35	1.63	2.96	3.29
						$\delta_\alpha = 0.5$						
98	1.61	3.00	4.37	0.86	3.09	5.57	2.02	2.97	3.36	2.03	2.98	3.36
196	2.18	3.05	3.89	0.82	3.08	4.96	1.58	2.97	3.31	1.52	2.96	3.29
						Panel D: $\delta_\beta = 3$						
						$\delta_\alpha = 0$						
98	2.05	3.10	4.13	1.04	3.12	4.77	1.41	2.95	3.53	1.34	2.98	3.44
196	2.22	3.05	3.92	0.96	3.12	4.85	1.13	2.97	3.37	1.08	2.96	3.43
						$\delta_\alpha = 0.1$						
98	1.97	3.07	4.19	0.88	3.06	4.98	1.22	2.94	3.51	1.02	2.95	3.45
196	2.23	3.07	4.04	0.99	3.14	4.81	0.91	2.97	3.36	1.14	2.97	3.34
						$\delta_\alpha = 0.3$						
98	1.88	3.08	4.30	1.14	3.12	5.31	1.35	2.95	3.47	1.56	2.96	3.41
196	2.18	3.05	3.94	0.91	3.10	4.88	1.05	2.96	3.34	1.21	2.95	3.34
						$\delta_\alpha = 0.5$						
98	1.52	3.01	4.48	0.52	3.10	5.63	1.68	2.97	3.40	1.82	2.97	3.38
196	2.13	3.08	4.03	1.13	3.09	5.45	1.21	2.96	3.29	1.04	2.96	3.30

Chapter 3

Inference in Threshold Regression after Model Selection

3.1 Introduction

The problem of model uncertainty is pervasive in economic applications because economic theory does not provide strong guidance about the variables to be included in the model mainly because the relevant theories are openended. That is, in many economic contexts there do not exist good theoretical reasons to include a particular set of theories or proxies a priori because these theories or proxies provide mutually compatible explanations of the underlying economic phenomenon. Brock and Durlauf (2001b) introduced the concept of openendedness who argued that this problem makes the coefficient estimates of interest fragile (Leamer (1978)). This means that inclusion or exclusion of a variable can result in substantial changes in magnitudes, loss of statistical significance, or, even switch signs. A standard approach to model uncertainty is to engage in model selection based on a post-single approach, that is, apply a conservative test or a selection criterion or a shrinkage and selection method in high-dimensional contexts (e.g., lasso) and then reestimate the model. However, these model selection methods do not explicitly address the impact of model uncertainty on inference and under certain cases they can give rise to misleading inference (e.g., Hansen (2005), Leeb and Potscher (2008), Potscher (2009)).

In this chapter, we are interested in making inference on threshold regression when there is uncertainty about the set of relevant regressors. A threshold regression model classifies observations into regime-specific models depending on whether the observed value of a threshold variable is above or below a threshold parameter. Usually, each regime obeys a linear model, although it can be extended to other nonlinear functions as well. Both the regression coefficients and the threshold parameter are estimated by the data. This model has been widely applied to various micro and macro contexts. For example, Durlauf and Johnson (1995) and Tan (2010) in cross-country growth behavior; Papageorgiou (2002) and Glushenkova, Kourtellos, and Zachariadis (2018) in trade; Durlauf, Kourtellos, and Tan (2017) in intergenerational mobility; Hansen (2017) in public debt. However, issue of model uncertainty was never been explicitly addressed in these studies.

Of course, the idea of incorporating model uncertainty within threshold regressions framework is not a new one. Kapetanios (1999) used information criteria to select the lag order in SETAR models. Lee, Seo, and Shin (2016) developed a LASSO estimator which both selects covariates and estimates the threshold parameter γ. These selection methods may work well when the coefficient on the nuisance or control variable is either "considerably away from zero" or "negligibly different from zero". However, when it is "moderately close to zero" then such post-single approaches can give rise to substantial size and power distortions in the threshold effect test.

We propose that the post-double selection method of Belloni, Chernozhukov, and Hansen (2011, 2014) (hereafter, referred to as BCH) can be applied in constructing a threshold test, which is valid under model uncertainty. BCH developed their method in the context of treatment effects. The ideas of post-double selection method is based on the partialling out technique of Frisch-Waugh-Lovell in the linear setting, the Neyman's $C(\alpha)$ test in the nonlinear setting (Neyman (1979)), and Robinson (1988) in the semi-parametric setting. In particular, we extend the post-double method to threshold regression, which can be viewed as a regression discontinuity model with an unknown discontinuity point (Yu (2014)). One challenge is that under the null of no threshold effects the threshold parameter (sample split value) is not identified. This makes the inference not standard even before one considers the issue of model uncertainty. Nevertheless, we show that the post-double procedure can still be successfully applied to construct moment conditions that are immune to model selection. In this way we address the issue of model uncertainty.

The chapter is organized as follows. Section 3.2 illustrates the problem of post-single approach in a simple threshold model and how the post-double approach can be applied to solve this issue. Section 3.3 presents our model and provides theoretical results of the post-double approach. Section 3.4 presents our Monte Carlo experiments. Section 3.5 concludes and discusses future work.

3.2 Implications of model selection

We begin our analysis by illustrating the problem of inference on threshold effects in the context of threshold regression when there is uncertainty about the set of regressors. To fix ideas we illustrate this problem in an idealistic environment of poverty traps where the entire family dynasty of child becomes trapped in a low-income regime when the parent's permanent income is below a certain threshold value conditional on some variables that captures family or neighborhood environment. Similarly, the entire family dynasty of child always remains in the high-income regime when parent's permanent income is above a threshold value. Such threshold type models are implied, for example, by the existence of borrowing constraints (e.g., Galor and Zeira (1993) and Han and Mulligan (2001)) or from the existence of social influences in models with strict stratification of neighborhoods by income (e.g., Benabou (1996), Durlauf (1996a), Durlauf and Seshadri (2018)). Our problem is that we wish to make inference on the presence of poverty trap but economic theory does not identify exactly the set of control variables. For instance, these variables may include proxies related to segregation, income distribution, local public finance institutions, early education, K-12 education, college education, local labor market, family structure, social capital.

In particular, let the dependent variable y_i denote the child's permanent income and let the threshold variable q_i denote the parent's permanent income. Consider a simple threshold model that allows for such poverty traps using different intercepts depending on whether parent's income is above or below an unknown threshold parameter γ

$$y_i = \theta_0 + aw_i + \delta_0 d_i(\gamma_0) + e_i, \quad i = 1, ..., n \tag{3.1}$$

where $d_i(\gamma_0) = 1\{q_i \leq \gamma_0\}$ is an indicator function that takes the value 1 if $q_i \leq \gamma_0$ and 0 otherwise and w_i is a scalar variable that captures the environment in which children develop. $\delta_0 = \zeta_0 - \theta_0$ is the difference between the regime specific intercepts. $e_i|(w_i, q_i)$ is $i.i.d$ from $N(0, \sigma_e^2)$ and $\{y_i, w_i, q_i\}_{i=1}^n$ is an $i.i.d$ sample from a DGP $\mathcal{P}_n$, where the means and variances of y_i, w_i, q_i are normalized to be zero and one, respectively.

In this chapter we are concerned with testing the null hypothesis $H_0 : \delta_0 = 0$ when there is uncertainty about the inclusion of w_i. The standard post-single-selection method for inference applies a model selection method to (3.1) and then reestimates the model and makes inference on δ_0 accordingly. Note that the threshold parameter of γ_0 can be estimated by a concentrated least squares method. Under certain assumptions the asymptotic distribution of the estimator of γ_0 involves two independent Brownian motions and the confidence intervals for γ_0 can be obtained by an inverted likelihood ratio approach (Hansen (2000)). The regression coefficients for the two regimes are then obtained using least-squares estimation on the two sub-samples, separately, with standard asymptotic theory. Under the null hypothesis of a linear model (i.e., no threshold effect), the threshold parameter, γ, is not identified, and hence inference is not standard. Hansen (1996) proposed a bootstrap procedure under which the bootstrap statistic approximates the asymptotic distribution and hence p-values constructed from the bootstrap are asymptotically valid assuming correct specification and no uncertainty about the inclusion of w_i. What happens to the properties of this bootstrap test when one first engages in the selection of w_i?

3.2.1 Post-single selection

We first explain the problem of the standard post selection method assuming γ_0 is known using standard omitted variable arguments as in BCH. Then, the argument trivially carries over to the case of estimated threshold parameter γ_0 since the bootstrap method of Hansen (1996) will be invalid. This is because the bootstrap supLR test will no longer be valid as it will based on an inconsistent conditional distribution function and hence it will no longer mimic well the asymptotic distribution of the t-test of δ_0.

Without loss of generality assume that γ_0 is known and

$$d_i(\gamma_0) = \kappa w_i + u_i \tag{3.2}$$

where $u_i|w_i$ is $i.i.d.$ from $N(0, \sigma_u^2)$ and u_i is independent of e_i. This also implies that $\sigma_d^2 = \kappa^2\sigma_w^2 + \sigma_u^2$. Finally, let $c_n = \frac{\sigma_e}{\sigma_w\sqrt{1-\rho^2}}$, where n is the sample size and $\rho = \kappa\frac{\sigma_w}{\sigma_d}$ is the correlation between $d_i(\gamma_0)$ and w_i. We assume that (3.1) and (3.2) hold for the collection of all DGPs $\mathcal{P}_n \in \mathcal{P}$.[1]

The nuisance variable w_i affects both directly the outcome variable y_i and indirectly via its effect on the threshold variable in equation (3.2). In the presence of uncertainty about the inclusion of w_i, standard post-single selection method will exclude w_i with probability 1 if

$$|a| \leq n^{-1/2} l_n c_n, \text{ for some } \; l_n \to \infty, \tag{3.3}$$

and include w_i with probability 1 if

$$|a| > n^{-1/2} l_n c_n, \text{ for some } \; l_n' > l_n, \tag{3.4}$$

where l_n and l_n' are slowly varying sequences. These sequences are important because they determine the behavior of the $\hat{\delta}(\gamma_0)$.

[1] We assume that all true parameter values and the DGP may change with the sample size n to capture the idea of a close to zero coefficient.

In particular, when a is large enough so that there exist sequences of models P_n such that (3.4) holds, w_i is included with probability one and standard inference based on asymptotic normality is valid. Similarly, when $a = o(1/\sqrt{n})$, and ρ is bounded away from one then w_i is excluded with probability one and again standard inference based on asymptotic normality is valid because the omitted variable bias due to the exclusion of regressor is small enough.

However, there are plausible sequences of DGPs where post-single selection generates substantial omitted variable bias. For example, consider the case when the coefficient on the uncertain regressor w_i is "moderately close to zero", that is, $a = \frac{l_n}{\sqrt{n}} c_n$. In this case, standard conservative inference (e.g., t-test) will exclude w_i with probability one. Note that

$$\frac{\sqrt{n}(\hat{\delta}(\gamma_0) - \delta_0)}{\sigma_n} = \underbrace{\frac{\sqrt{n} E_n(d_i(\gamma_0) e_i)}{\sigma_n E_n d_i(\gamma_0)^2}}_{(i)} + \underbrace{\frac{\sqrt{n} E_n(d_i(\gamma_0) w_i) a}{\sigma_n E_n d_i(\gamma_0)^2}}_{(ii)} \tag{3.5}$$

where $\sigma_n^2 = \sigma_e^2 / \sigma_d^2$. While term $(i) \xrightarrow{d} N(0,1)$, $|(ii)| \geq \frac{1}{2} \frac{|\rho|}{1-\rho^2} l_n \propto l_n \to \infty$. This implies that the standard post-selection estimator is neither asymptotically normal nor $\sqrt{n}$-consistent

$$\left| \frac{\sqrt{n}(\hat{\delta}(\gamma_0) - \delta_0)}{\sigma_n} \right| \to \infty.$$

3.2.2 Post-double selection

In this chapter we show that a post-double procedure proposed by BCH can be adapted to provide robust inference for the threshold effect in the sense that the threshold test is not sensitive to inclusion or exclusion of the control variable w_i. As discussed in the introduction the post-double procedure is based on the idea of partialling out technique of Frisch-Waugh-Lovell theorem.

To see this, first consider the case of known γ_0. The post double procedure works as follows:

Step 1: Select to include w_i if it is a significant regressor of y_i based on a conservative t-test in model (3.1) .

Step 2: Select to include w_i if it is a significant regressor of $d_i(\gamma_0)$ based on a conservative t-test in model (3.2) .

Step 3: If w_i is selected in at least one of steps 1 and 2 use model (3.1) otherwise use model

$$y_i = \theta + \delta d_i(\gamma_0) + e_i \tag{3.6}$$

to make inference on δ.

Therefore, the post double procedure implies that we exclude w_i with probability 1 if both

$$|a| < \frac{l'_n}{\sqrt{n}} c_n \text{ and } |\kappa| < \frac{l'_n}{\sqrt{n}} (\sigma_u/\sigma_w), \tag{3.7}$$

which means that to exclude w_i we require both $|a| = O_p(\frac{1}{\sqrt{n}})$ and $|\kappa| = O_p(\frac{1}{\sqrt{n}})$. So when w_i is excluded, the omitted variable bias is negligible $|ii| \leq 2\frac{l'^2_n}{\sqrt{n}} \to 0$, and hence the post-double estimator $\tilde{\delta}(\gamma_0)$ is consistent and asymptotically normal $\frac{\sqrt{n}(\tilde{\delta}(\gamma_0) - \delta_0)}{\sigma_n} \to N(0,1)$, as $n \to \infty$ regardless of whether w_i is included or excluded.

In other words, the post-double procedure constructs moment conditions for the parameter of interest δ_0 that are robust to misspecification, in the sense that they have an immunization condition. That is, we can estimate the empirical analog of the moment condition $M(\delta_0, a) = 0$, where δ_0 is the parameter of interest and a is the nuisance parameter and enjoy the property that

$$\frac{dM(\delta_0, a)}{da} = 0, \tag{3.8}$$

which means the moments are locally unaffected by minor perturbations of the nuisance parameter around the true parameter values.

We can obtain more intuition by illustrating the above immunization property in the context of model (3.1). In doing so, note that the above procedure can be viewed as an application of the Frisch-Waugh-Lovell theorem.

1. Regress y_i on w_i and obtain residuals $\tilde{u}^y$.
2. Regress $d_i(\gamma_0)$ on w_i and obtain the residuals $\tilde{u}^d(\gamma_0)$.
3. Regress $\tilde{u}^y$ on $\tilde{u}^d(\gamma_0)$.

Then, for the known γ_0 the resulting moment condition is

$$E((\tilde{u}^y - \tilde{u}^d(\gamma_0)\delta_0)\tilde{u}^d(\gamma_0)) = 0 \tag{3.9}$$

which is the empirical analog of
$M(\delta_0, a) = E([(y_i - E(y_i|w_i)) - (d_i(\gamma_0) - E(d_i(\gamma_0)|w_i)\delta_0)][d_i(\gamma) - E(d_i(\gamma_0)|w_i)])$, for which the derivative of the moment condition is free of a

$$\frac{dE([(y_i - E(y_i|w_i)) - (d_i(\gamma_0) - E(d_i(\gamma_0)|w_i)\delta_0)][d_i(\gamma_0) - E(d_i(\gamma_0)|w_i)])}{da}|_{a=a_0} = 0. \tag{3.10}$$

Hence, when estimation and inference is based on the empirical analog of moments $M(\delta_0, a)$, post double procedure implies that $\frac{dM(\delta_0,a)}{da} = 0$. In this sense, inference on δ is immune from model uncertainty.

In threshold regressions, however, the threshold parameter γ is unknown and it is estimated. In this case, we do not have $d_i(\gamma)$ but $\hat{d}_i = d_i(\hat{\gamma})$ and the moment condition (3.9) is replaced by $E([(y_i - E(y_i|w_i)) - (\hat{d}_i - E(\hat{d}_i|w_i)\delta_0)][\hat{d}_i - E(\hat{d}_i|w_i)]$. If we show that $E(\hat{d}_i - E(\hat{d}_i|w_i)) \xrightarrow{p} E(d_i(\gamma) - E(d_i(\gamma)|w_i))$, then the condition (3.8) will hold in the case of the threshold regressions framework as well. However, we do not know if there is a threshold effect or not. In the case of a threshold effect then $E(\hat{d}_i - E(\hat{d}_i|w_i)) \xrightarrow{p} E(d_i(\gamma_0) - E(d_i(\gamma_0)|w_i))$, while in a linear model $E(\hat{d}_i - E(\hat{d}_i|w_i)) \xrightarrow{p} d_i(\gamma^*) - E(d_i(\gamma^*)|w_i)$, where γ^* is a random variable.

In the next section, we show that the threshold parameter estimator obtained by a linear model indeed converges in distribution to a random variable γ^* and that the above immunization property carries over to case of estimated γ. Therefore, the results of post-double selection are naturally extended to threshold regressions. Next, we proceed to formally define the threshold regression model and the post-double procedure.

3.3 The threshold model

We start by generalizing model (3.1) to include the focus $(k-1) \times 1$ vector of regressors x_i in addition to the doubtful scalar regressor w_i

$$y_i = \begin{cases} a_1 w_i + \vartheta_1' x_i + e_i, & q_i \leq \gamma, \\ a_2 w_i + \vartheta_2' x_i + e_i, & q_i > \gamma, \end{cases} \tag{3.11}$$

where q_i can be part of x_i or the same as w_i. It is useful to rewrite model (3.11) as a single equation. Define $h_i = [x_i, w_i]'$, $h_i(\gamma) = [x_i 1\{q_i \leq \gamma\}, w_i 1\{q_i \leq \gamma\}]'$, and $\theta = [\vartheta, a]'$, and $\delta = [\delta_1, \delta_2]'$ then

$$y_i = \theta' h_i + \delta h_i(\gamma) + e_i \tag{3.12}$$

where $E(e_i|h_i, q_i) = 0$.

In our case the parameters of interest are ϑ_1, ϑ_2 (or ϑ and δ_1 equivalently) and the nuisance parameters are a and δ_2. Hence,we are interested in testing

$$H_0 : \delta_1 = 0 \text{ vs. } H_1 : \delta_1 \neq 0 \tag{3.13}$$

when one engages in selection of the doubtful regressor w_i. It is important to point out that the uncertainty about the inclusion of w_i may affect the presence of threshold effects, since imperfect model selection may lead to substantial omitted variable bias, which in turn will result to inconsistent conditional distribution function.

3.3.1 Post-double procedure

Consider the auxiliary system of $k-1$ regression equations

$$x_i(\gamma) = W_i'\Pi + U_i, \tag{3.14}$$

where $E(U_i|W_i) = 0$ and $W_i = diag\{w_i, ..., w_i\}$.[2]

Step 1: Select to include w_i if it is a significant regressor of y_i based on a conservative joint Wald test in model (3.11), i.e.,

$$H_0 : a_1 = a_2 = 0 \text{ vs. } H_1 : \text{Not } H_0 \tag{3.15}$$

Step 2: Select to include w_i if it is a significant regressor of $x_i(\hat{\gamma})$ based on a conservative joint Wald test in the SUR model (3.14), i.e.,

$$H_0 : \Pi = 0 \text{ vs. } H_1 : \Pi \neq 0, \tag{3.16}$$

Step 3: If w_i is selected in at least one of steps 1 and 2 use model (3.12) otherwise use model

$$y_i = \begin{cases} \varphi_1' x_i + e_i, & q_i \leq \lambda, \\ \varphi_2' x_i + e_i, & q_i > \lambda, \end{cases} \tag{3.17}$$

to make inference on δ_1.

3.3.2 Inference in Step 1

In the first step test, we perform a test for the inclusion of w_i in our model, which will be later tested for threshold effects. Consequently, we do not know a priori if we have a linear or a threshold model and we do not want to put any extra restrictions on ϑ_1 and ϑ_2.

Therefore it is important that the inference on w_i should be valid regardless of whether the true model is linear or threshold model. We are going to show that the inference

[2] Equivalently, we can estimate the system equation-by-equation.

on w_i in step 1, is valid both in the presence and absence of threshold effects in the rest parameters. The following assumption is necessary to derive the asymptotic properties of Step 1 test. Let $M(\gamma) = E(h_i h_i' 1\{q_i \leq \gamma\})$, $D(\gamma) = E(h_i h_i' | q_i = \gamma)$ and $V(\gamma) = E(h_i h_i' e_i^2 | q_i = \gamma)$ and $f(q)$ the density function of q_i,γ_0 the true value of γ, $D = D(\gamma_0)$, $V = V(\gamma_0)$, $f = f(\gamma_0)$ and $M = E(h_i h_i')$.

Assumption 1

1.1 (h_i, q_i, e_i) is $i.i.d.$

1.2 $E(e_i | F_{i-1}) = 0$.

1.3 $E|h_i|^{2r} < \infty$ and $E|h_i e_i|^{2r} < \infty$.

1.4 For all $\gamma \in \Gamma$, $E(|h_i|^{2r} | q_i = \gamma) \leq C$ and $E(|h_i|^{2r} e_i^{2r} | q_i = \gamma) \leq C$ for some $C < \infty$, $f(\gamma) \leq \bar{f} < \infty$.

1.5 $f(\gamma)$, $D(\gamma)$, $V(\gamma)$ are continuous at $\gamma = \gamma_0$.

1.6 $\delta_n = c_n^{-\alpha}$ with $c \neq 0$, $0 < \alpha < \frac{1}{2}$.

1.7 $c' D c > 0$, $c' V c > 0$ and $f > 0$.

1.8 $M > M(\gamma) > 0$ for all $\gamma \in \Gamma$.

Assumption (1.1) is an assumption for the dependence. Assumption (1.2) guarantees the correct specification of the conditional mean. Assumptions (1.3) and (1.4) are unconditional and conditional moment bounds. Assumption (1.5) requires the threshold variable to have a continuous distribution, and the conditional variance to be continuous at γ, which excludes regime-dependent heteroskedasticity. Assumption (1.6) states that the difference in regression slopes gets small as the sample size increases (we are taking an asymptotic approximation valid for small values of δ_n) and allows us to reduce the rate of convergence, in order to derive a simpler asymptotic distribution. Assumption (1.7) is a full-rank condition needed to have nondegenerate asymptotic distributions. Assumption(1.8) is a conventional full-rank condition which excludes multicollinearity.

Since we are agnostic about the presence of threshold effects, we need to keep in mind that the asymptotic behavior of $\hat{\gamma}$ will depend on whether the true model is linear or threshold. The following Proposition describes the asymptotic distribution in each case.

Proposition 3.1 *Under Assumption 1, $H_0 : a_1 = a_2 = 0$ and $E(e_i)^4 < \kappa < \infty$ we have:*

1. $n^{1-2\alpha}(\hat{\gamma}-\gamma_0) \xrightarrow{d} \omega T$ *when* $\vartheta_1 \neq \vartheta_2$.

2. $\hat{\gamma} \xrightarrow{d} \gamma^* = F_q^{-1}(\lambda^*)$ *when* $\vartheta_1 = \vartheta_2$.

where $\omega = \frac{C'VC}{(C'DC)^2 f}$, $T = arg\max_{-\infty<r<\infty}[-\frac{1}{2}r + W(r)]$, *F(.) denoting the distribution function of* q_i *and* $I(q_i \leq \gamma) \equiv I(F(q_i) \leq \lambda)$, $\hat{\lambda} \xrightarrow{d} \lambda^*$ *and* $\lambda^* = arg\max_{\lambda\in[\underline{\lambda},\overline{\lambda}]} B_u'(1,\lambda)B_u(1,\lambda) + (E_{xx'}^{1/2}B_u(1) - E_{xx',\gamma}^{1/2}B_u(1,\lambda)'(E_{xx'} - E_{xx',\gamma})^{-1}(E_{xx'}^{1/2}B_u(1) - E_{xx',\gamma}^{1/2}B_u(1,\lambda))$ *and* $E_{xx} = E(x_i'x_i)$ *and* $E_{xx,\gamma} = E(x_i'x_i 1\{q_i \leq \gamma\})$.

The first result is the standard result in threshold regressions as it was shown in Hansen (2000). The second result, shows that the threshold parameter estimated while the true model is linear, converges in distribution to an argument that maximizes a linear combination of squares of Brownian bridges.

Next, we construct the Wald statistic and derive its asymptotic distribution under the two alternative cases:

$$W_n(\hat{\gamma}) = \frac{(R'\hat{\beta})'(R'(X(\hat{\gamma})'X(\hat{\gamma}))^{-1}R)^{-1}(R'\hat{\beta})}{s^2} \tag{3.18}$$

where R is a selection matrix, $\hat{\gamma}$ is the γ estimated under the null, s^2 is the residual variance calculated under the alternative, $\hat{\beta} = [\vartheta, a, \delta_1, \delta_2]$ and $X(\gamma)$ is the matrix with stacked elements $[x_i 1\{q_i \leq \gamma\}, w_i 1\{q_i \leq \gamma\}, x_i 1\{q_i > \gamma\}, w_i 1\{q_i > \gamma\}]$.

The following Proposition, shows us that irrespectively of the presence of threshold effects or not, the Wald statistic converges in distribution to a chi-squared distribution.

Proposition 3.2 *Under Assumption 1,* $H_0 : a_1 = a_2 = 0$, $E(e_i)^4 < \kappa < \infty$ *and* $\alpha < \frac{1}{4}$*:*

$$W_n(\hat{\gamma}) \xrightarrow{d} \chi_2^2 \tag{3.19}$$

holds if

1. $\vartheta_1 \neq \vartheta_2$.
2. $\vartheta_1 = \vartheta_2$ *and* e_i *is independent of* x_i.

Hence, the Wald test is valid and the BCH method carries over to threshold regressions framework.

In the next section, we assess the performance of the Wald statistic in finite samples. Furthermore, we perform two Monte Carlo experiments to highlight the implications of the post single selection method on the size of the test, and the performance of the proposed post double selection method.

3.4 Monte Carlo

In this section we show that the standard post-single selection can have adverse consequences in the bootstrap threshold test approach of Hansen (1996) both in terms of size and power. We also provide evidence that a post-double approach restores the properties of the bootstrap threshold test.

Consider the following data generating process which represents a linear model

$$\begin{aligned} y_i &= a_1 w_i + \vartheta_1' x_i + e_i, q_i \leq \gamma \\ y_i &= a_2 w_i + \vartheta_2' x_i + e_i, q_i > \gamma \end{aligned} \tag{3.20}$$

and

$$q_i = c w_i + v_i \tag{3.21}$$

where $v_{1i} \sim N(0, 0.01)$, $x_i = [1, x_{1i}]$, $x_{1i} \sim N(0,1)$, $w_i \sim N(0,1)$, $e_i \sim N(0, 0.25)$, $\vartheta_1 = \vartheta_2 = (2,2)'$, $a_1 = a_2 = 0$, $c = 0.8$.

Furthermore, consider the data generating process that represents a threshold regression model

$$\begin{aligned} y_i &= a_1 w_i + \vartheta_1' x_i + e_i, q_i \leq \gamma \\ y_i &= a_2 w_i + \vartheta_2' x_i + e_i, q_i > \gamma \end{aligned} \tag{3.22}$$

and

$$q_i = c w_i + v_i \tag{3.23}$$

where $v_{1i} \sim N(0, 0.01)$,$x_i = [1, x_{1i}]$, $x_{1i} \sim N(0,1)$, $w_i \sim N(0,1)$, $e_i \sim N(0, 0.25)$, $\vartheta_1 = (3,3)$,$\vartheta_2 = (2,2)'$, $a_1 = a_2 = 0$, $c = 0.8$.

Table (3.1) shows the size of the test of the null hypothesis $H_0 : a_1 = a_2 = 0$ for different sample sizes (n=100,250, 500). The number of simulation draws were 1000. We can see that in both cases of a linear and a threshold model, the size of the test is very close to 0.05. We, further assess the power of the test by varying a. Table (3.1) displays also the results of this exercise. As the magnitude of a increases, the power of the tests becomes larger and approaches 1 in both linear and threshold regression.

Note that a changes with the sample size so we can not comment about the power as the sample size increases because a decreases.

Having shown that the first step step is valid without the need of knowledge for the presence or not of threshold effects, we now return to the post single selection method and its implications on the size of the test.

We are considering two different data generating processes in order to justify our concerns about typical post selection procedure. We are trying to capture cases in which the regressor for which we face uncertainty is correlated with the regressors that they should be included and the threshold varible in order to create omitted variable bias of a significant amount that will affect the inference for threshold effects. In the first data generating process, the regressor of question is both correlated with threshold effects and the regressor that should be included, while in the second is correlated ony with the threshold effect. We use 1000 simulations and consider three different sample sizes $n = 100, 250, 500$.

For the first DGP, we face uncertainty about the inclusion of threshold variable q_i as a regressor, which is also correlated with x_{1i}

$$y_i = 2 + 2x_{1i} + aw_i + (\delta_o + \delta_1 x_{1i} + \delta_2 w_i)(1\{\gamma\}) + e_i$$
$$x_{1i} = cw_i + v_i$$

where $e_i \sim N(0,1)$, $v_i \sim N(0,1)$, $q_i \sim N(0,1)$ and $w_i = q_i$.

We control the degree of correlation between w_i and x_{1i} by varying c. Furthermore, we vary a, the coefficient of w_i in DGP. [3] In the left panel of Table (3.2), the size of the test after post single procedure is presented. As a increases the distortion of the size increases, because the omitted variable bias increases. As c increases, hence the correlation with the included regressor x_{1i} increases, the distortion of the size decreases, since the effect of w_i is captured from x_{1i}.

In our second DGP, w_i is correlated with q_i through the equation $q_i = cw_i + v_i$, hence the omitted variable bias will occur due to the correlation with the threshold effect

$$y_i = a_1 w_i + \vartheta_1' x_i + e_i, q_i \leq \gamma$$
$$y_i = a_2 w_i + \vartheta_2' x_i + e_i, q_i > \gamma$$

[3] We tried several values of a and we present those that created distortion of the size. Higher values, for example $a = \frac{7}{\sqrt{n}}$ did not create distortion of the size because their effect was large enough to be chosen from the first step size, while cases with lower values such as $a = \frac{1}{4\sqrt{n}}$ did not create substantial bias that would affect inference.

and

$$q_i = cw_i + v_i$$

where $v_i \sim N(0, 0.01)$, $x_{1i} \sim N(0, 1)$, $w_i \sim N(0, 1)$, $e_i \sim N(0, 0.25)$.

We vary a and c as before. The results are displayed in the left panel of Table (3.2). It is worth noting that when $c = 0$ we do not observe any distortion of the size. This is the case where the regressor for which we face uncertainty is not correlated with none of the regressors we are interested in conducting inference and hence no omitted variable bias arises. Overall, we can see that the distortion of size is higher for higher values of c, since the omitted variable bias increases. As a increases we observe increase and then decrease probably due to the fact that $\frac{4}{\sqrt{n}}$ is in some times chosen due to its magnitude.

We have shown that under model uncertainty, the standard model selection method may lead to large omitted variable bias, which in turn will affect the size of the threshold test.

We have assesed the size of the test through the previous Monte Carlo simulations. The results for DGP 1 and DGP 2 are presented in the right panels of Table (3.2). For both DGP's, the distortion of the size is minimized and in some cases even eliminated. Overall, the results seem to be better for DGP 1. The reason is that w_i is correlated with both x_{1i} and the threshold effect through q_i, hence it is more possible to be chosen in the second step, thus less possible for omitted variable bias to arise.

We assess the performance of the method in terms of power as well. The results are displayed in Table (3.3) for each data generating process. When the threshold effect increases, the power of the test increases as expected. Morever for a given value of the threshold effect, the power of the test increases as the sample size increases, as expected.

3.5 Conclusion and Future Work

This chapter contributes to the literature of threshold regression by proposing a threshold test which is valid in the presence of uncertainty about the set of relevant regressors. We show that standard post-single selection practices can have adverse consequences in the bootstrap threshold test approach of Hansen (1996) both in terms of size and power. Following Belloni, Chernozhukov, and Hansen (2011, 2014), we propose a post-double selection that constructs moment conditions for the threshold effect that are robust to misspecification, in the sense that they have an immunization condition. Finally, a Monte Carlo simulation illustrates that post-double selection restores the size and the power of the bootstrap threshold supWald test.

In terms of future work, we believe that our proposed paper is immediately applicable to a wide range of questions with broad policy significance such as in intergenerational mobility, child development literature, and cross-country growth studies as discussed in the introducation. Methodologically, a natural extension of this paper is to consider a high dimensional regression model with a possible threshold along the lines of Lee, Seo, and Shin (2016) who propose a method of estimating and selecting relevant control variables from a set of many potential covariates. Another fruitful avenue to deal with model uncertainty is to generalize existing model averaging methods that apply to linear models (e.g., Brock and Durlauf (2001b), Raftery, Madigan, and Hoeting (1997), Hansen (2007, 2010)) to account for threshold effects.

3.6 Tables

Table 3.1: **First step test - Size and power**

This table shows the size (a=0) of the test of the null hypothesis $H_0 : a_1 = a_2 = 0$ and the power of the test ($a = 1/\sqrt{n}, 2/\sqrt{n}, 4/\sqrt{n}$)for different sample sizes (n=100,250, 500). The left panel corresponds to a linear model and the right panel to a threshold model.

	Linear Model				*Threshold Model*			
a	0	$1/\sqrt{n}$	$2/\sqrt{n}$	$4/\sqrt{n}$	0	$1/\sqrt{n}$	$2/\sqrt{n}$	$4/\sqrt{n}$
sample size								
100	0.04	0.19	0.65	1	0.05	0.55	0.86	0.98
250	0.04	0.17	0.66	1	0.06	0.59	0.91	1
500	0.04	0.16	0.65	1	0.04	0.59	0.92	1

Table 3.2: **Size of the test** $H_o : \vartheta_1 = \vartheta_2$

This table shows the size of the test of the null hypothesis $H_o : \vartheta_1 = \vartheta_2$. Panel A displays the size of the test for *DGP* 1 and Panel B for *DGP* 2, for different sample sizes (n=100,250, 500). Each panel is divided into two subpanels. The left subpanel corresponds to the post-single method and the right subpanel to the post-double method. We control the degree of correlation between w_i and x_{1i} by varying c. a is the coefficient of w_i in DGP.

	Post-Single				*Post-Double*			
	Values of c				*Values of c*			
	0	0.2	0.8	1	0	0.2	0.8	1
Values of a								
Panel A: *DGP 1*								
n=100								
$1/\sqrt{n}$	0.10	0.09	0.07	0.06	0.06	0.05	0.04	0.05
$2/\sqrt{n}$	0.27	0.28	0.16	0.13	0.06	0.06	0.04	0.05
$4/\sqrt{n}$	0.70	0.68	0.46	0.38	0.09	0.08	0.04	0.05
n=250								
$1/\sqrt{n}$	0.10	0.09	0.07	0.06	0.05	0.05	0.04	0.04
$2/\sqrt{n}$	0.30	0.28	0.16	0.14	0.06	0.05	0.04	0.04
$4/\sqrt{n}$	0.73	0.72	0.51	0.41	0.09	0.06	0.04	0.04
n=500								
$1/\sqrt{n}$	0.10	0.10	0.08	0.06	0.07	0.06	0.06	0.06
$2/\sqrt{n}$	0.29	0.28	0.17	0.13	0.07	0.06	0.06	0.06
$4/\sqrt{n}$	0.75	0.72	0.50	0.41	0.10	0.07	0.06	0.06
Panel B: *DGP 2*								
n=100								
$1/\sqrt{n}$	0.06	0.22	0.28	0.28	0.03	0.07	0.07	0.08
$2/\sqrt{n}$	0.06	0.47	0.71	0.71	0.03	0.08	0.10	0.12
$4/\sqrt{n}$	0.05	0.07	0.12	0.13	0.03	0.05	0.07	0.07
n=250								
$1/\sqrt{n}$	0.05	0.24	0.27	0.28	0.04	0.04	0.06	0.06
$2/\sqrt{n}$	0.05	0.48	0.72	0.72	0.04	0.06	0.10	0.09
$4/\sqrt{n}$	0.04	0.06	0.10	0.12	0.04	0.04	0.05	0.05
n=500								
$1/\sqrt{n}$	0.05	0.25	0.31	0.30	0.05	0.05	0.07	0.05
$2/\sqrt{n}$	0.05	0.49	0.73	0.73	0.05	0.07	0.08	0.07
$4/\sqrt{n}$	0.05	0.07	0.11	0.11	0.05	0.05	0.06	0.05

Table 3.3: **Power of the test-Post double method**

This table shows the power of the post double procedure. Panel A displays the power for *DGP* 1 and Panel B for *DGP* 2, for different sample sizes (n=100,250, 500). Each panel is divided into two subpanels. The left subpanel corresponds to $\delta_0 = \delta_1 = 0.5$, $\delta_2 = 0.05$ and the right subpanel to $\delta_0 = \delta_1 = 1$, $\delta_2 = 0.05$. a is the coefficient of w_i in DGP.

	Values of c				*Values of c*			
	0	0.2	0.8	1	0	0.2	0.8	1
Values of a								
				Panel A: *DGP 1*				
				n=100				
$1/\sqrt{n}$	0.136	0.13	0.125	0.132	0.27	0.28	0.34	0.38
$2/\sqrt{n}$	0.116	0.115	0.125	0.132	0.26	0.26	0.34	0.38
$4/\sqrt{n}$	0.118	0.115	0.125	0.132	0.24	0.25	0.34	0.38
				n=250				
$1/\sqrt{n}$	0.27	0.23	0.30	0.32	0.47	0.48	0.75	0.81
$2/\sqrt{n}$	0.25	0.22	0.30	0.32	0.47	0.48	0.75	0.81
$4/\sqrt{n}$	0.22	0.21	0.30	0.32	0.42	0.47	0.75	0.81
				n=500				
$1/\sqrt{n}$	0.34	0.32	0.52	0.58	0.67	0.73	0.95	0.97
$2/\sqrt{n}$	0.33	0.32	0.52	0.58	0.66	0.73	0.95	0.97
$4/\sqrt{n}$	0.29	0.31	0.52	0.58	0.64	0.72	0.95	0.97
				Panel B: *DGP 2*				
				n=100				
$1/\sqrt{n}$	0.55	0.34	0.26	0.27	0.66	0.43	0.40	0.40
$2/\sqrt{n}$	0.47	0.30	0.22	0.23	0.60	0.39	0.37	0.37
$4/\sqrt{n}$	0.47	0.27	0.19	0.20	0.60	0.37	0.34	0.33
				n=250				
$1/\sqrt{n}$	0.64	0.45	0.46	0.47	0.72	0.57	0.70	0.71
$2/\sqrt{n}$	0.60	0.40	0.42	0.43	0.69	0.53	0.68	0.68
$4/\sqrt{n}$	0.60	0.39	0.40	0.41	0.69	0.53	0.67	0.67
				n=500				
$1/\sqrt{n}$	0.72	0.55	0.66	0.66	0.74	0.71	0.88	0.88
$2/\sqrt{n}$	0.69	0.53	0.64	0.64	0.72	0.70	0.88	0.88
$4/\sqrt{n}$	0.69	0.52	0.63	0.63	0.72	0.69	0.87	0.87

Conclusions

The present dissertation aims in providing three complementary approaches to the study of intergenerational mobility. We contribute to the literature of intergenerational mobility by proposing novel econometric methods in order to address limitations in existing methodologies.

In the first chapter, we focus on the timing of parental investments during childhood and young adulthood. Surprisingly, much of the literature has focused on the linear model and the IGE coefficient, ignoring timing which is a very important aspect of parental investments. Using functional regressions we treat the observations as "snapshots" of an underlying latent curve, where annual income data are treated as discrete signals of a latent income process. Hence, the object of interest is a curve that captures the intergenerational trajectory of an individual. We find that parental investments during early and late childhood or young adulthood are generally more productive than middle childhood. These findings indicate that income shocks play a crucial role in parental human capital investments in children and in their long run outcomes. We further investigate the heterogeneity of the trajectories with respect to parental income, parental education, and family structure. Timing of the shocks related to socioeconomic status and family structure plays a key role in intergenerational trajectories of individuals, especially for disadvantaged children.

In Chapter 2, a general threshold spatial autoregression model is proposed which nests several models including the spatial autoregression model and spatial autoregression-mixed regression model. Our framework allows for both fixed and diminishing threshold effects and we develop a two-step GMM estimation method that exploit both linear and quadratic moment conditions. We study the limiting properties of the estimators of the threshold parameter and slope parameters of spatial lags and regression coefficients. Exploiting the smoothness of the GMM criterion, we reduce the rate of convergence of the threshold parameter. As a result, not only the estimators of the slope parameters of spatial lags and regression coefficients, but the estimator of the threshold parameter as well, are normally distributed. We assess the performance of the proposed estimation method using a Monte Carlo simulation.

Chapter 3 contributes to the literature of threshold regression by proposing a threshold test which is valid under misspecification, in the sense that we face uncertainty about the set of relevant regressors. We show the implications of the standard post-single selection practices on the size of the bootstrap threshold test approach of Hansen (1996). Adopting the method of Belloni, Chernozhukov, and Hansen (2011, 2014) to threshold regressions, we propose a post-double selection that constructs moment conditions for the threshold effect that are robust to misspecification. Monte Carlo Simulations show that post-double selection restores the size and the power of the bootstrap threshold supWald test.

Bibliography

Abbott, B., and G. Gallipoli, 2019, Permanent-Income Inequality,, University of British Columbia, Working Paper.

Almond, D., and J. Currie, 2011, Killing Me Softly: The Fetal Origins Hypothesis, *Journal of Economic Perspectives* 32, 153—172.

Andreou, E., E. Ghysels, and A. Kourtellos, 2010, Regression Models With Mixed Sampling Frequencies, *Journal of Econometrics* 158, 246–261.

Anselin, L., 1980, Estimation methods for Spatial Autoregressive Structures, *Regional Science Dissertation and Monograph Series* 8, 273.

——— , 1988, *Spatial Econometrics: Methods and Models.* (Springer).

Becker, G. S., S. D. Kominers, K. M. Murphy, and J. L. Spenkuch, 2018, A Theory of Intergenerational Mobility, *Journal of Political Economy* 126, S7–S25.

Becker, G. S., and N. Tomes, 1979, An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility, *Journal of Political Economy* 87, 1153–1189.

——— , 1986, Human Capital and the Rise and Fall of Families, *Journal of Labour Economics* 4, 1–39.

Belloni, A., V. Chernozhukov, and C. Hansen, 2011, Inference for High-Dimensional Sparse Econometric Models, *Advances in Economics and Econometrics: Tenth World Congress Volume 3, Econometrics.*

——— , 2014, Inference on Treatment Effects after Selection among High-Dimensional Controls, *Review of Economic Studies* 81, 608–650.

Benabou, R., 1996, Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance, *American Economic Review* 86, 584–609.

Benhabib, J., A. Bisin, and M. Jackson, 2011a, Handbook of Social Interaction, *Journal of the European Economic Association* 1.

——— , 2011b, Handbook of Social Interaction, *Journal of the European Economic Association* 2.

Bernal, R., 2008, The Effect of Maternal Employment and Child Care On Children's Cognitive Development, *International Economic Review* 49, 1173–1209.

——— , and M. P. Keane, 2010, Quasi-structural Estimation of a Model of Childcare Choices and Child Cognitive Ability Production, *Journal of Econometrics* 156, 164 – 189.

——— , 2011, Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers, *Journal of Labor economics* 29, 459—512.

Blundell, B. R., L. Pistaferri, and I. Preston, 2008, Consumption Inequality and Partial Insurance,, *American Economic Review* 126, 1887–1921.

Brock, W. A., and S. N. Durlauf, 2001a, Discrete Choice with Social Interactions, *Review of Economic Studies* 1993, 235—260.

Brock, W. A., and S. N Durlauf, 2001b, Growth Empirics and Reality, *World Bank Economic Review* 15, 229–272.

Brock, W. A., and S. N. Durlauf, 2002, A Multinomial Choice Model with Neighborhood Effects, *American Economic Review Papers and Proceedings* 92, 298–303.

Caetano, G., J. Kinsler, and H. Teng, 2017, Towards Causal Estimates of Children's Time Allocation on Skill Development, University of Rochester Working Paper.

Card, D., and L. Giuliano, 2013, Peer Effects and Multiple Equilibria in the Risky Behavior of Friends, *The Review of Economics and Statistics* 95, 1130–1149.

Card, D., Mas A., and J. Rothstein, 2008, Tipping and the Dynamics of Segregation., *The Quarterly Journal of Economics* 123, 177–218.

Cardot, H., F. Ferraty, and P. Sarda, 2003, Spline Estimators for the Functional Linear Model, *Statistica Sinica* 13, 571–591.

Carneiro, P., and J. J. Heckman, 2003, Human Capital Policy, in J. J. Heckman, A. B. Krueger, and B. M. Friedman, ed.: *Inequality in America: What Role for Human Capital Policies?* vol. 4 . pp. 77–239 (MIT Press: Cambridge).

Carneiro, P., L. G. Italo, K. Salvanes, and E. Tominey, 2018, Intergenerational Mobility and the Timing of Parental Income, *IZA Discussion Paper Series* 9479.

Carneiro, P., K. G. Salvanes, and E. Tominey, 2016, Family Income Shocks and Adolescent Human Capital,, University of York, Working Paper.

Caucutt, E. M, and L. Lochner, 2017, Early and Late Human Capital Investments, Borrowing Constraints, and the Family, Western University, Working Paper No. 2017-3.

Chan, K. S., 1993, Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *The Annals of Statistics* 21, 520–533.

Chausse, P., T. Chen, and K. A. Couch, 2015, Assessing the Inter-generation Mobility – A Functional Data Regression Approach, .

Cliff, A.D., and J.K. Ord, 1973, Spatial Autocorrelation, *Progress in Human Geography* 19,2, 245–249.

Conley, T., and G. Topa, 2001, Socio-Economic Distance and Spatial Patterns in Unemployment, *Journal of Applied Econometrics* 17, 303–327.

Cooper, R., and A. John, 1988, Coordinating Coordination Failures in Keynesian Models, *The Quarterly Journal of Economics* 103, 441–463.

Cunha, F., and J. J. Heckman, 2007, The Technology of Skill Formation, *American Economic Review* 97, 31–47.

——— , and S. M. Schennach, 2010, Estimating the Technology of Cognitive and Noncognitive Skill Formation, *Econometrica* 78, 883–931.

de Paula, A., 2009, Inference in a Synchronization Game with Social Interactions, *Journal of Econometrics* 148, 56–71.

Del Boca, D., C. Flinn, and M. Wiswall, 2014, Household Choices and Child Development, *The Review of Economic Studies* 81, 137–185.

——— , 2016, Transfers to Families with Children and Child Outcomes,, *Economic Journal* 526, F138–F183.

Durlauf, S. N., 1993, Nonergodic Economic Growth,, *Review of Economic Studies* 60, 349–366.

——— , 1996a, A Theory of Persistent Income Inequality, *Journal of Economic Growth* 1, 75–93.

——— , 1996b, Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality, in W. Barnett, G. Gandolfo, and C. Hillinger, ed.: *Dynamic Disequilibrium Modelling: Proceedings of the Ninth International Symposium on Economic Theory and Econometrics,* (Cambridge University Press: Cambridge).

——— , and Y. M. Ioannides, 2010, Social Interactions, *Annual Review of Economics* 2, 451–478.

Durlauf, S. N., and P. Johnson, 1995, Multiple Regimes and Cross-Country Growth Behavior, *Journal of Applied Econometrics* 10, 365–384.

Durlauf, S. N., A. Kourtellos, and C. M. Tan, 2017, Status Traps, *Journal of Business and Economic Statistics* 35, 265–287.

Durlauf, S. N., and A. Seshadri, 2018, Understanding the Great Gatsby Curve, *NBER Macroeconomics Annual* 32, 333–393.

Friedman, M., 1957, The Permanent Income Hypothesis, in *A Theory of The Consumption Function,* . pp. 20–37 (Princeton University Press).

Galor, O., and J. Zeira, 1993, Income Distribution and Macroeconomics, *Review of Economic Studies* 60, 35–52.

Glaeser, E., B. Sacerdote, and J. Scheinkman, 1996, Crime and Social Interactions, *Quarterly Journal of Economics* 111, 507—548.

Glushenkova, M., A. Kourtellos, and M. Zachariadis, 2018, Barriers to Price Convergence, *Journal of Applied Econometrics* 33, 1081–1097.

Gonzalo, J., and J.-Y. Pitarakis, 2002, Estimation and Model Selection Based Inference in Single and Multiple Threshold Models, *Journal of Econometrics* 110, 319–352.

Gupta, A., 2018, Nonparametric Specification Testing via the Trinity of Tests., *Journal of Econometrics* 203, 169–185.

Hai, R., and J. J. Heckman, 2017, Inequality in Human Capital and Endogenous Credit Constraints,, *Review of Economic Dynamics* 25, 4–36.

Haider, S., and G. Solon, 2006, Life-Cycle Variation in the Association between Current and Lifetime Earnings, *American Economic Review* 96, 1308–1320.

Han, S., and C. Mulligan, 2001, Human Capital, Heterogeneity and Estimated Degrees of Intergenerational Mobility, *Economic Journal* 111, 207–243.

Hansen, B. E., 1996, Inference when a Nuisance parameter is Not Identified Under the Null Hypothesis, *Econometrica* 64, 413–430.

——— , 2000, Sample Splitting and Threshold Estimation, *Econometrica* 68, 575–603.

——— , 2005, Challenges for Econometric Model Selection, *Econometric Theory* 21, 60–68.

——— , 2007, Least Squares Model Averaging, *Econometrica* 75, 1175–1189.

——— , 2010, Averaging Estimators for Autoregressions with a Near Unit Root, *Journal of Econometrics* 158, 142–155.

——— , 2017, Regression Kink With an Unknown Threshold, *Journal of Business and Economic Statistics* 35, 228–240.

——— , and M. Caner, 2001, Threshold Autoregression With A Unit Root, *Econometrica* 69, 1555–1596.

Hansen, K., J. J. Heckman, and K. J. Mullen, 2004, The Effect of Schooling and Ability on Achievement Test Scores, *Journal of Econometrics* 121, 39 – 98.

Heckman, J., and S. Mosso, 2014, The Economics of Human Development and Social Mobility, *Annual Review of Economics* 6, 689–733.

Kapetanios, G., 1999, Model Selection in Threshold Models., Cambridge Working Papers in Economics 9906.

Keane, M. P., and K. I. Woplin, 2001, The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment, *International Economic Review* 42, 1051–1103.

Kelejian, H. H., and I. R. Prucha, 1998, A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a spatial Autoregressive Model with Autoregressive Disturbances, *The Journal of Real Estate Finance and Economics* 17, 99–121.

Landers, R., and J. Heckman, 2016, The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S., *The Scandinavian Journal of Economics* 119, 178–230.

Leamer, E. E., 1978, *Specification Searches* (Wiley, New York.).

Lee, L.-F., 2007, GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, *Journal of Econometrics* 137, 489–514.

Lee, S., MH. Seo, and Y. Shin, 2016, The Lasso for High Dimensional Regression with a Possible Change Point, *Journal of the Royal Statistical Society: Series B* 78, 193–210.

Leeb, H., and B. M. Potscher, 2008, Can One Estimate the Unconditional Distribution of Post-Model-Selection Estimators?, *Econometric Theory* 24, 338–376.

Lin, X., and L. Lee, 2010, GMM Estimation of Spatial Autoregressive Models With Heteroskedasticity, *Journal of Econometrics* 157, 34–52.

Loury, G., 1981, Intergenerational Transfers and the Distribution of Earnings, *Econometrica* 49, 843–867.

Malikov, E., and Y. Sun, 2017, Semiparametric Estimation and Testing of Smooth Coefficient Spatial Autoregressive Models, *Journal of Econometrics* 199, 12–34.

Manski, Charles F., 1993, Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies* 60, 531—542.

Mazumder, B., 2005, Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data, *Review of Economics and Statistics* 87, 235–255.

——— , 2015, Estimating the Intergenerational Elasticity and Rank Association in the U.S.: Overcoming the Current Limitations of Tax Data, *Working Paper Series WP-2015-4, Federal Reserve Bank of Chicago.*

Moon, S. H., 2014, Multi-dimensional Human Skill Formation with Multi-dimensional Parental Investment, University of Chicago, Working Paper.

Morris, J. S., 2015, Functional Regression, *Annual Review of Statistics and Its Application* 2, 321–359.

Newey, W., and D. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, 4, 2113–2241.

Neyman, J., 1979, C(a) Tests and Their Use., *Sankhya* 41, 1–21.

Pan, J., 2015, Gender Segregation in Occupations: The Role of Tipping and Social Interactions, *Journal of Labor Economics* 33, 365 – 408.

Papageorgiou, C., 2002, Trade as a Threshold Variable for Multiple Regimes, *Economics Letters* 77, 85–91.

Potscher, B. M., 2009, Confidence Sets Based on Sparse Estimators Are Necessarily Large, *Sankhyā: The Indian Journal of Statistics, Series A (2008-)* 71, 1–18.

Raftery, A. E., D. Madigan, and J. A. Hoeting, 1997, Bayesian Model Averaging for Linear Regression Models, *Journal of the American Statistical Association* 92, 179–191.

Ramsay, J., and B. Silverman, 1997, Functional Data Analysis, .

Robinson, P., 1988, Root- N-Consistent Semiparametric Regression, *Econometrica* 56, 931–54.

Seber, G., 2008, A Matrix Handbook for Statisticians, *Wiley.*

Seo, M. H., and Y. Shin, 2016, Dynamic Panels with Threshold Effect and Endogeneity, *Journal of Econometrics* 195, 169–186.

Sirakaya, S., 2006, Recidivism and Social Interactions, *Journal of the American Statistical Association* 101, 863—-877.

Solon, G., 1992, Intergenerational Income Mobility in the United States, *American Economic Review* 82, 393—-407.

Su, l., and S. Jin, 2010, Profile Quasi-maximum Likelihood Estimation of Partially Linear Spatial Autoregressive Models, *Journal of Econometrics* 157, 18–33.

Tan, C. M., 2010, No One True Path: Uncovering the Interplay Between Geography, Institutions, and Fractionalization in Economic Development, *Journal of Applied Econometrics* 25, 1100–1127.

Taspinar, S., Dogan O., and W. P. M. Vijverberg, 2018, GMM Inference in Spatial Autoregressive Models., *Econometric Reviews* 37, 931–954.

Topa, G., 2001, Social Interactions, Local Spillovers and Unemployment, *Review of Economic Studies* 68, 261–95.

Wang, J.-L., J.-M. Chiou, and H.-G. Muller, 2016, Functional Data Analysis, *Annual Review of Statistics and Its Application* 3, 257–295.

Wu, Y., Y. Fan, and H-G. Muller, 2010, Varying-coefficient Functional Linear Regression, *Bernoulli* 16, 730–758.

Yao, F., Y. Fu, and T. C. M. Lee, 2011, Functional Mixture Regression, *Biostatistics* 12, 341–353.

Yao, F., and H.-G. Muller, 2010, Functional Quadratic Regression, *Biometrika* 97, 49–64.

Yu, P., 2014, Understanding Estimators of Treatment Effects in Regression Discontinuity Designs, *Econometric Reviews.*

Zhang, X., and J. L. Wang, 2015, Varying-coefficient Additive Models for Functional Data, *Biometrika* 102, 15–32.

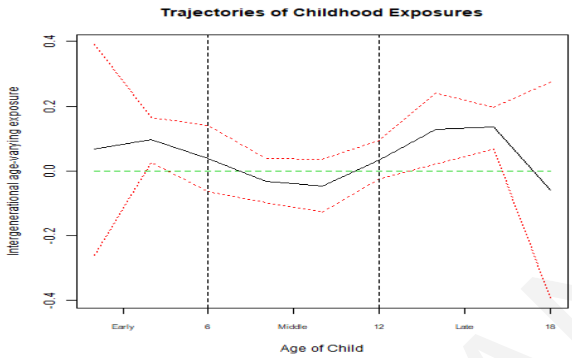
Appendices

Appendix A

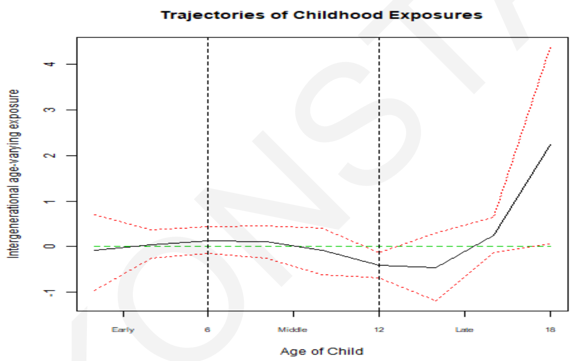
Figure A1: **Intergenerational Trajectories of Bi-annual Income (short sample)**

This figure presents the estimates of models (1.5) and (1.10) for the bi-annual short sample.

(a) Bi-annual Income



(b) Bi-annual Stock of Income



(c) Bi-annual Partial Effects

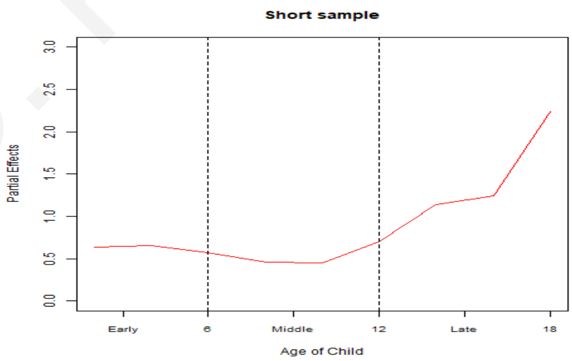
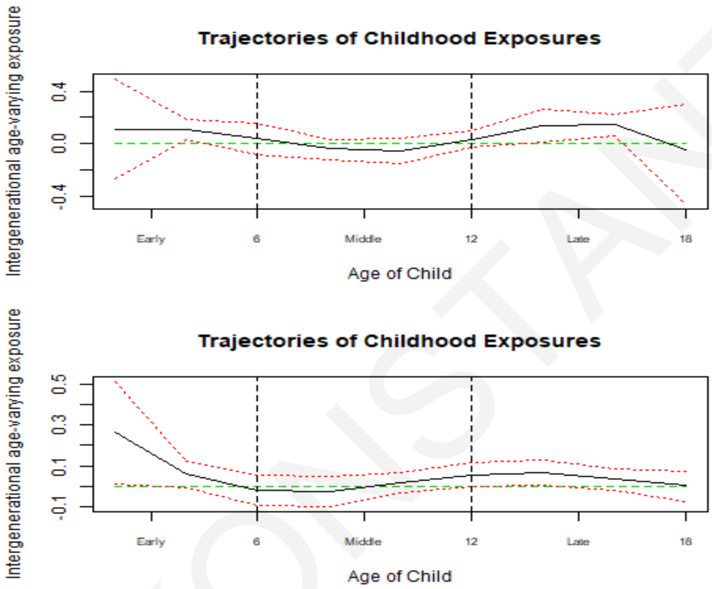


Figure A2: **Restricted Bi-annual Sample**

This figure presents results restricting the bi-annual sample to include same individuals in both short and long samples. A2(a) displays the estimates from model (1.5) and A2(b) from model (1.10). The upper panel of each sub-figure correspons to short sample and the lower to the long sample.

(a) Bi-annual Income



(b) Bi-annual Stock of Income

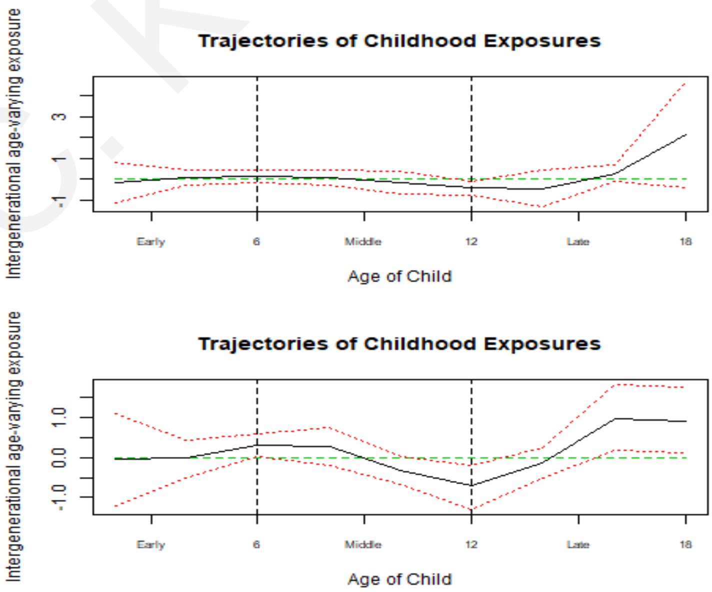


Figure A3: **Restricted Sample**

This figure presents the partial effects of stock of income for daughters and sons, when the short sample is restricted to have the same number of individuals as the long.

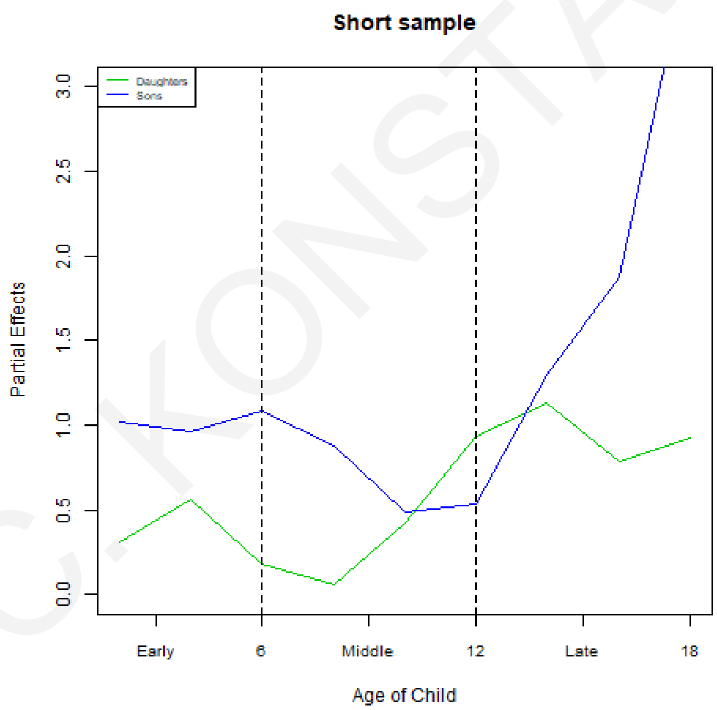
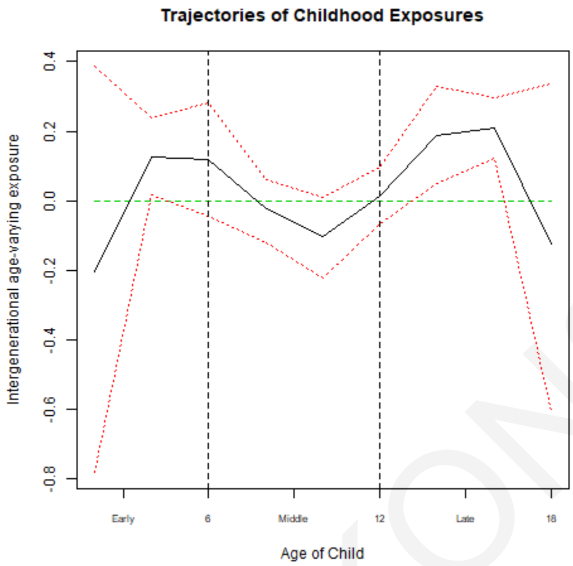


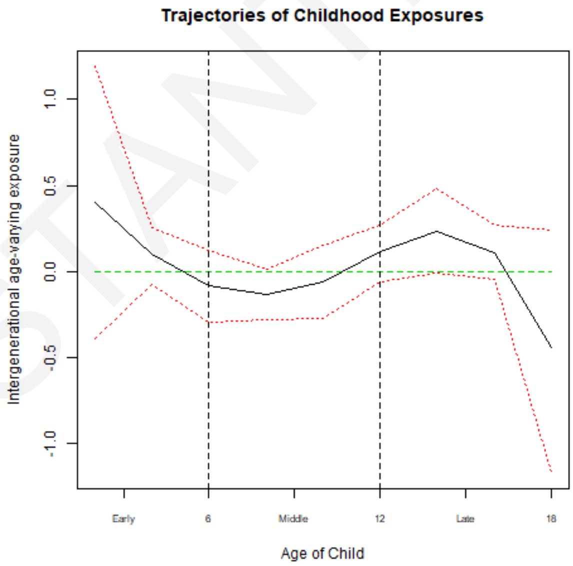
Figure A4: **Intergenerational Trajectories of Annual Income by Parental Income Quartiles (short sample)**

This figure presents the baseline results from model (1.5) based on the short sample. Figures A4(a), (b), (c), (d) present the estimates of intergenerational elasticity function $\hat{\beta}(t)$ for the first parent's permanent income quartile, second, third and fourth quartile respectively. The red dotted lines represent the bootstrap confidence bands.

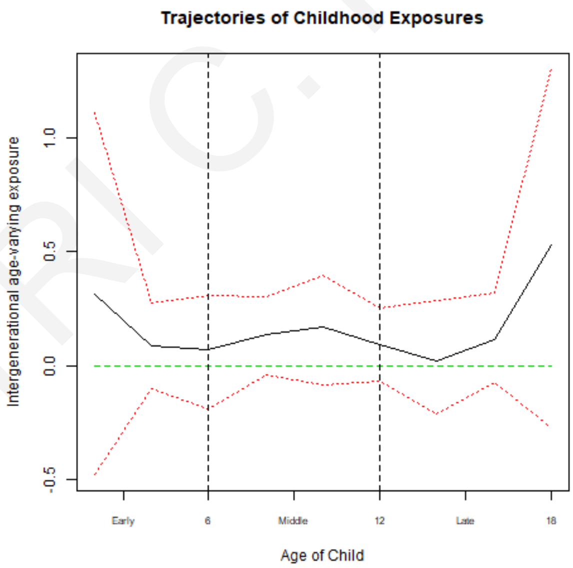
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

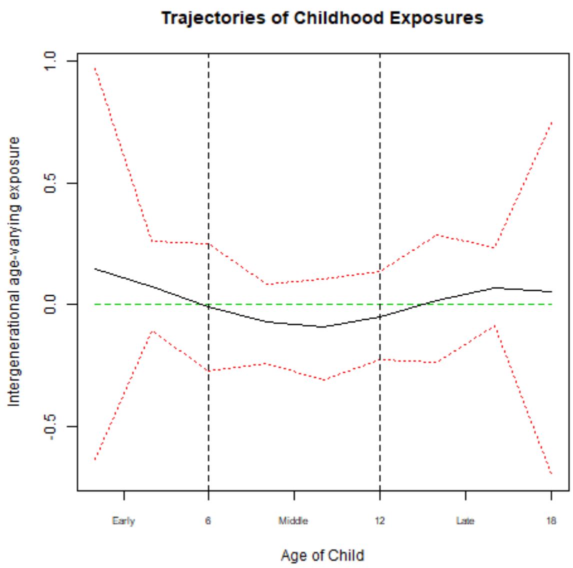
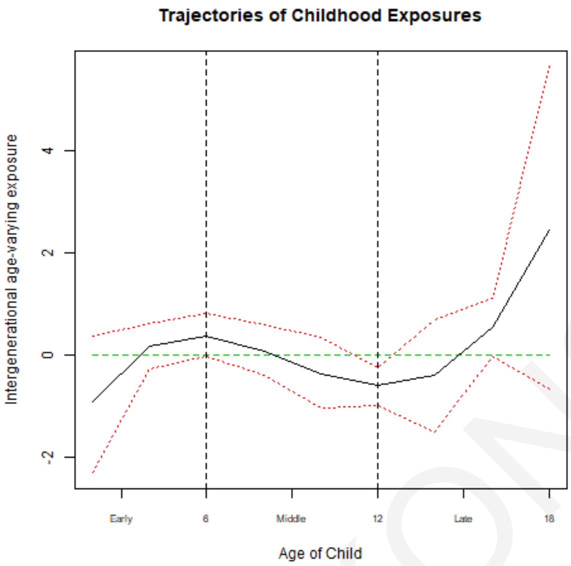


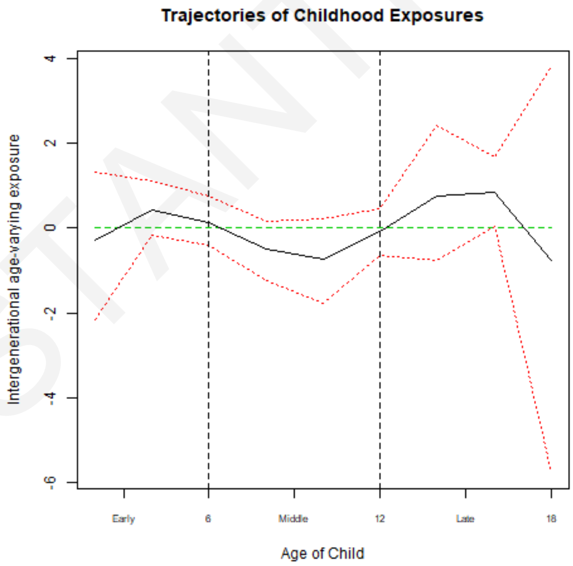
Figure A5: **Intergenerational Trajectories of Stock of Income by Parental Income (short sample)**

This figure presents the baseline results from model (1.10) based on the short sample. Figures A5(a), (b), (c), (d) present the estimates for the first parent's permanent income quartile, second, third and fourth quartile respectively. The red dotted lines represent the bootstrap confidence bands.

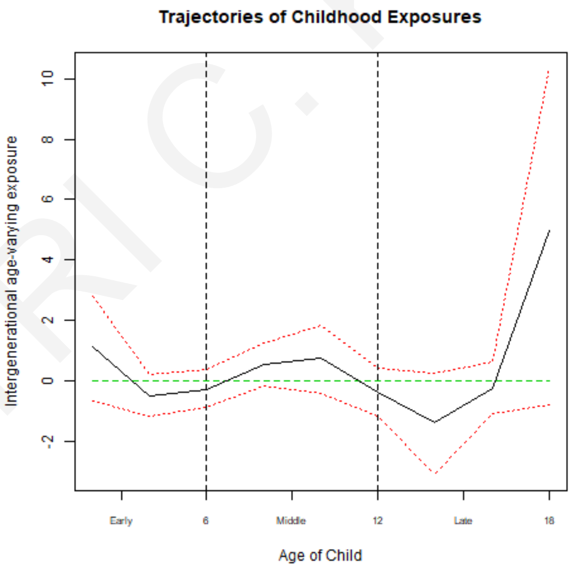
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

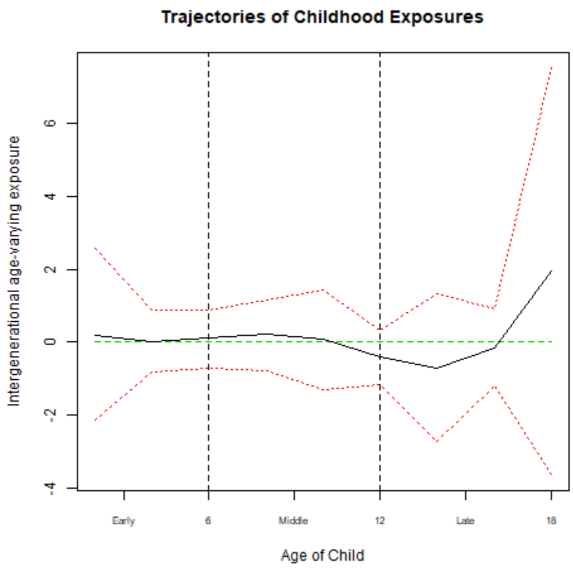


Figure A6: **Partial Effects of Stock of Income by Parental Income Quartiles**

This figure presents the trajectory partial effects of equation (1.10) based on the short sample. The red line corresponds to first parental permanent income quartile, the green line to the second quartile, the blue line to third quartile and the cyan line to fourth quartile.

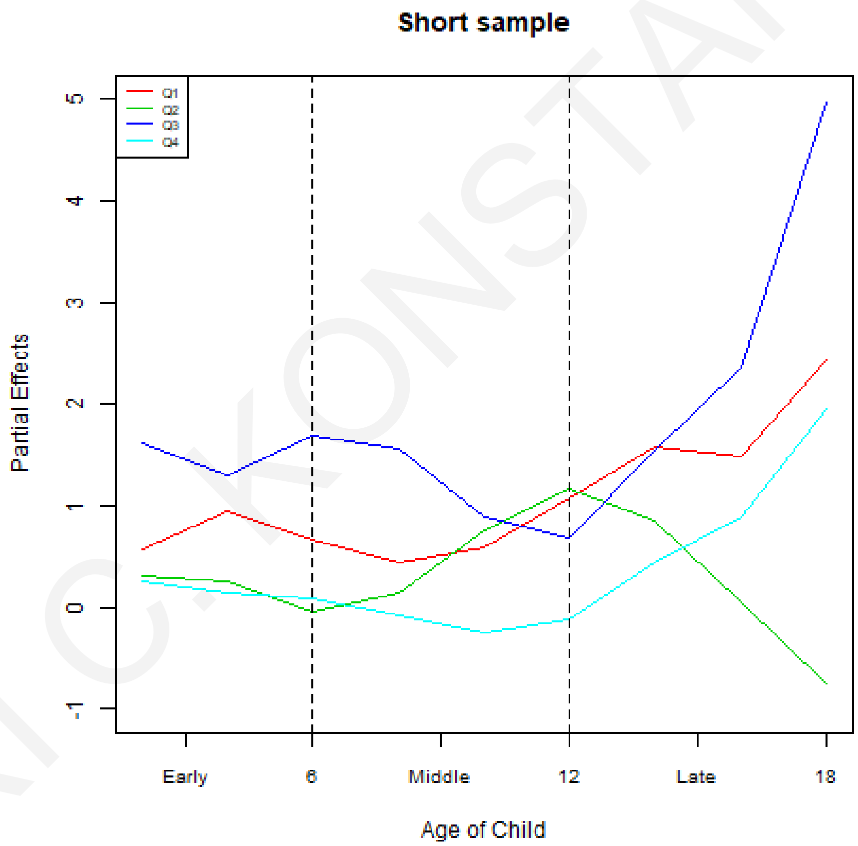
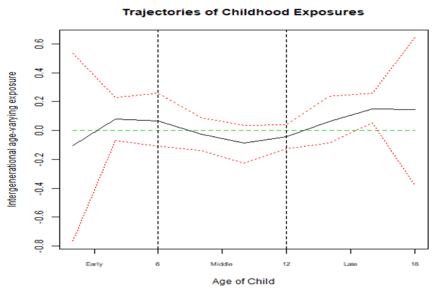


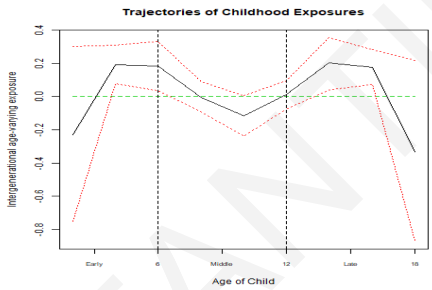
Figure A7: **Timing of Income Shocks - Trajectories of Income (short sample)**

This figure compares the intergenerational trajectories of income for the short sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

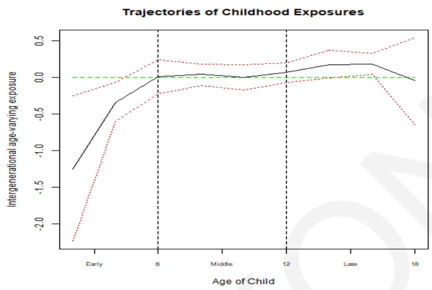
(a) First Quartile - Early



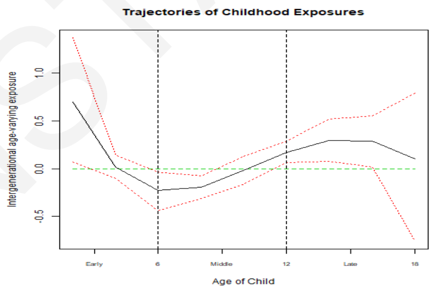
(b) First Quartile - Late



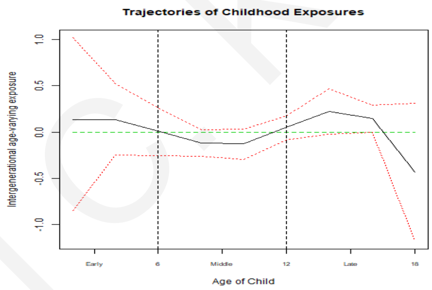
(c) Second Quartile - Early



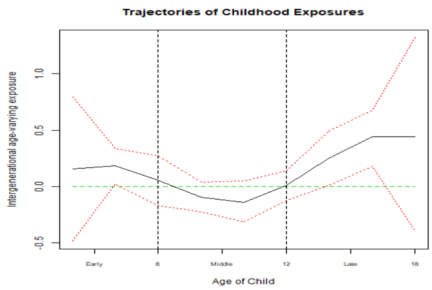
(d) Second Quartile - Late



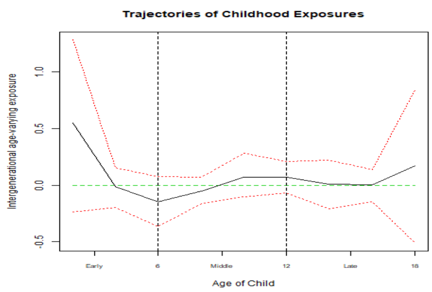
(e) Third Quartile - Early



(f) Third Quartile - Late



(g) Fourth Quartile - Early



(h) Fourth Quartile - Late

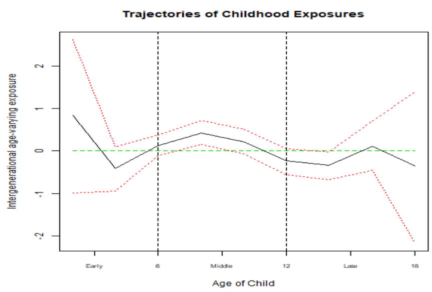
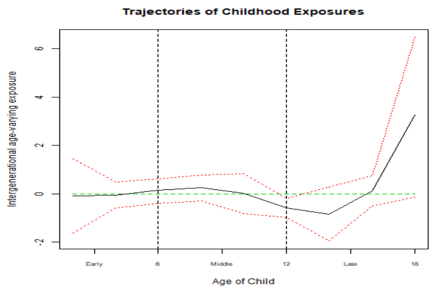


Figure A8: **Timing of Income Shocks - Trajectories of Stock of Income (short sample)**

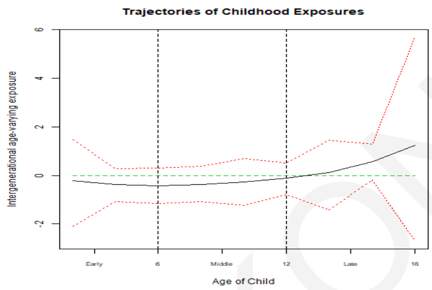
This figure compares the intergenerational trajectories of marginal effects for stocks of income for the short sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

(a) First Quartile - Early

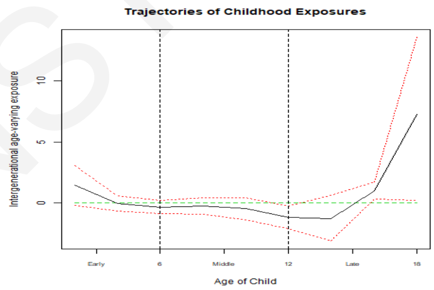


(b) First Quartile - Late

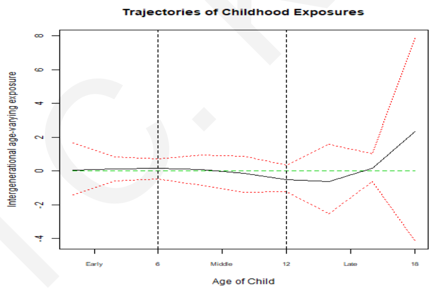
(c) Second Quartile - Early



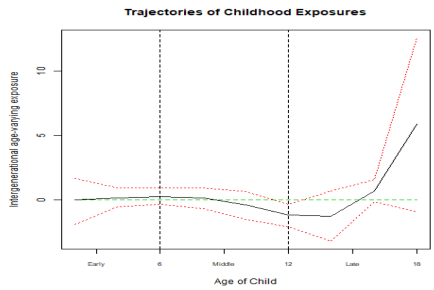
(d) Second Quartile - Late



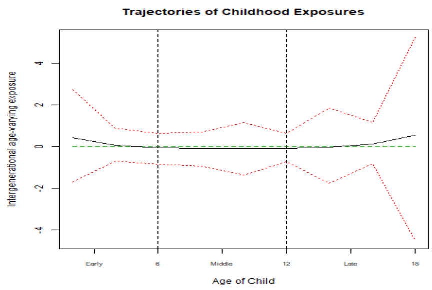
(e) Third Quartile - Early



(f) Third Quartile - Late



(g) Fourth Quartile - Early



(h) Fourth Quartile - Late

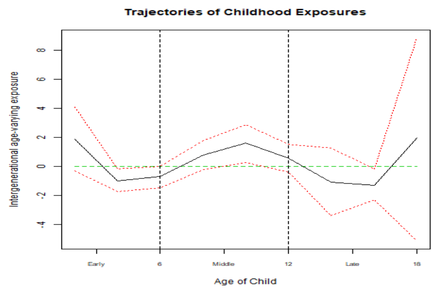
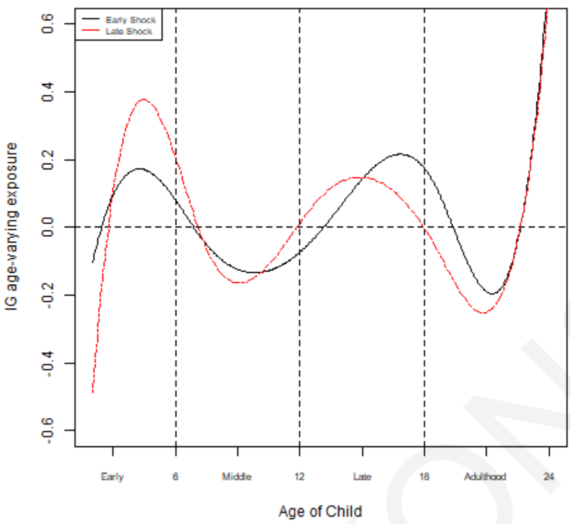


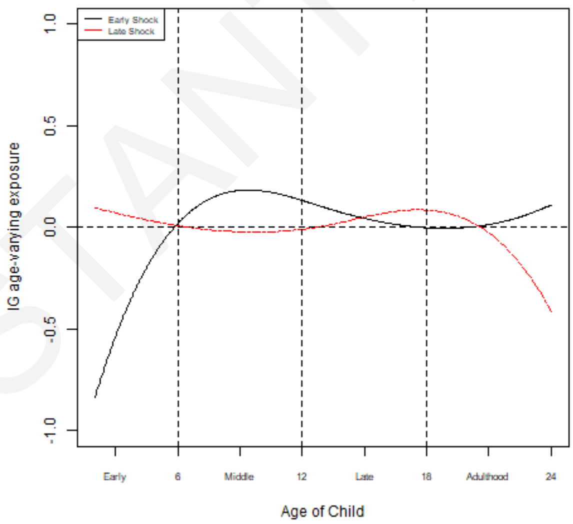
Figure A9: **Timing of Income Shocks - Trajectories of Income**

This figure compares the intergenerational trajectories of income for the long sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

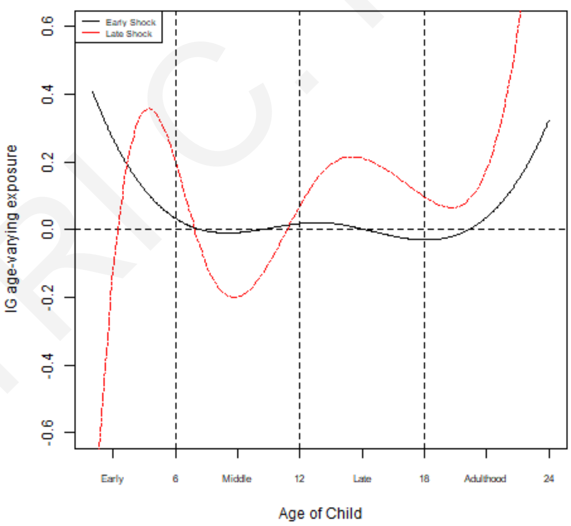
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(c) Fourth Quartile

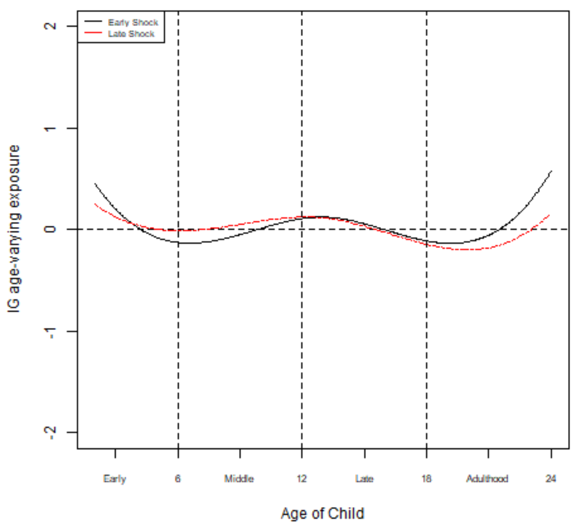
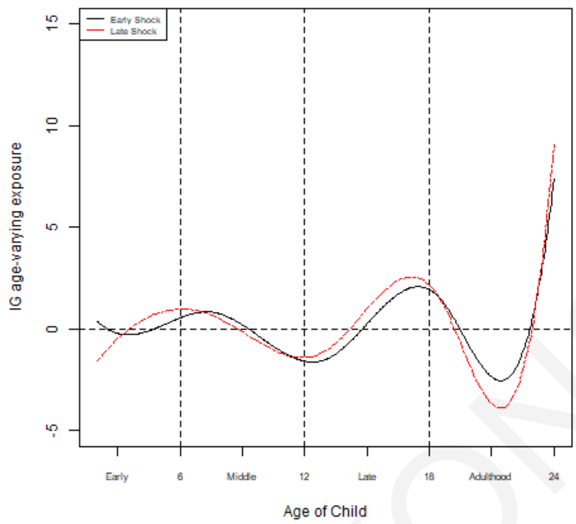


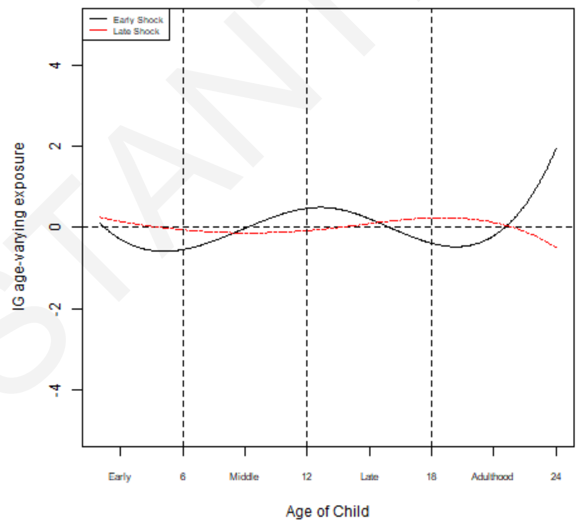
Figure A10: **Timing of Income Shocks - Trajectories of the Stock of Income**

This figure compares the intergenerational trajectories of marginal effects for stocks of income for the long sample based on quartiles of parent's income during early childhood against the ones based on quartiles of parent's income during late childhood.

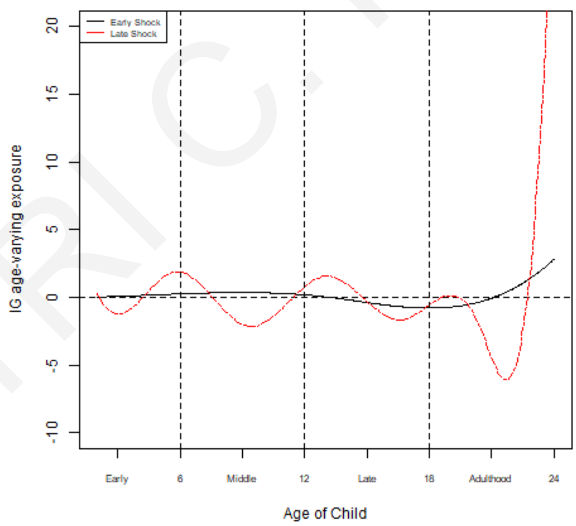
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(c) Fourth Quartile

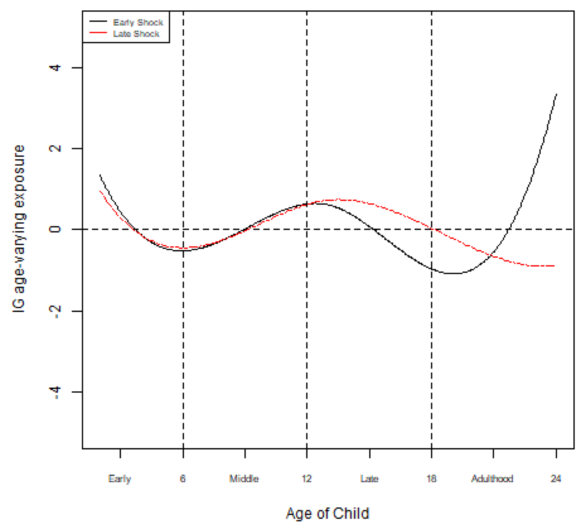
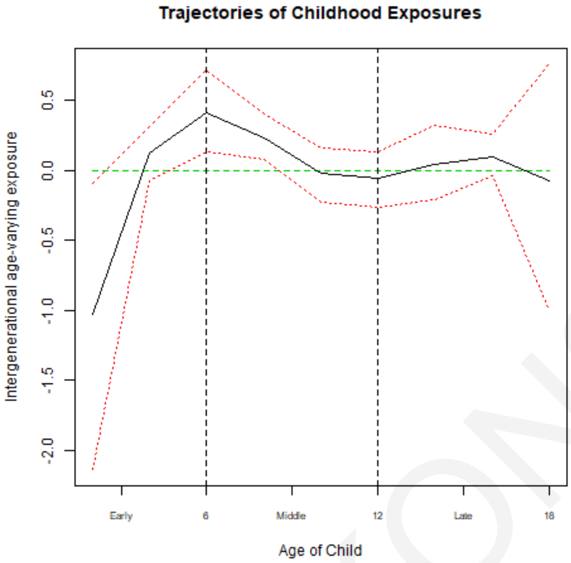


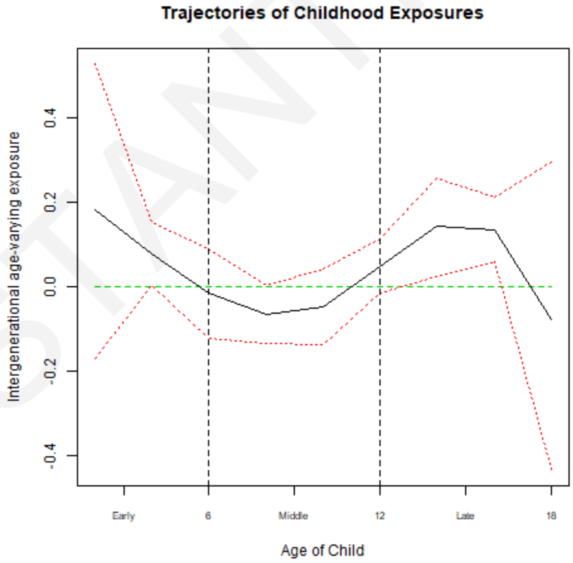
Figure A11: **Intergenerational Trajectories of Annual Income by Father's Education**

This figure presents the results from model (1.5) based on the short sample for the sub-samples based on father's education.

(a) Non-high School Graduates



(b) High School Graduates



(c) Non-college Graduates

(d) College Graduates

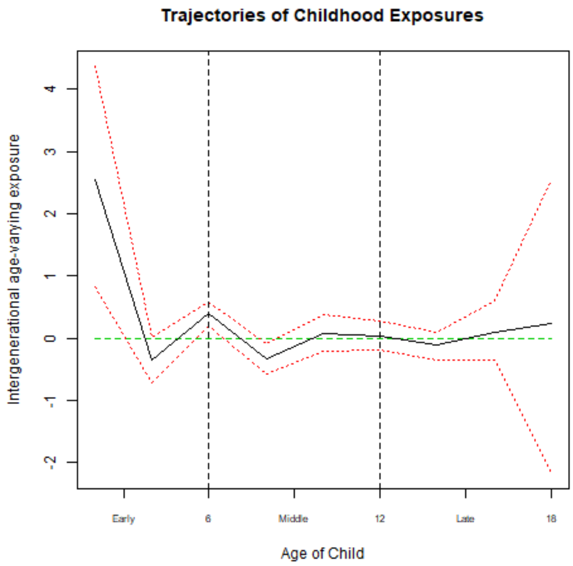
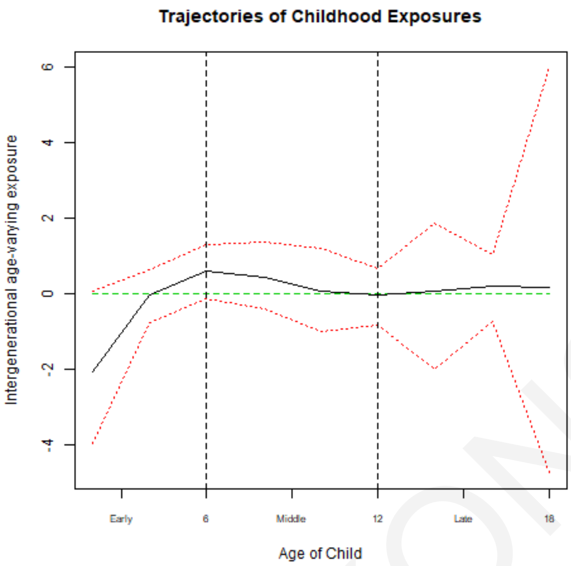


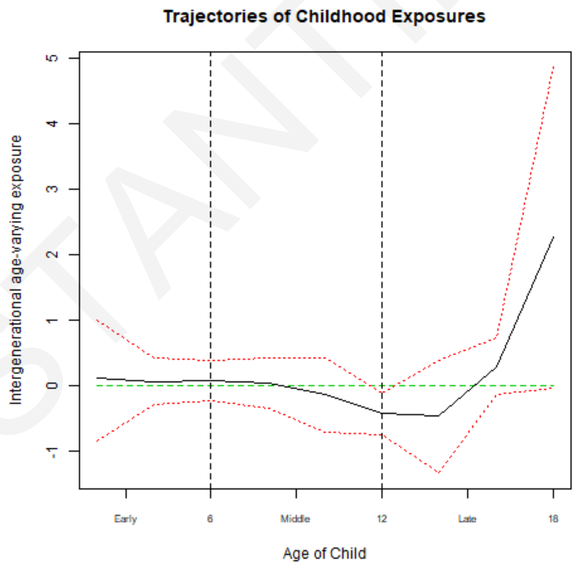
Figure A12: **Intergenerational Trajectories of Stock of Income by Father's Education**

This figure presents the results from model (1.10) based on the short sample for sub-samples based on father's education.

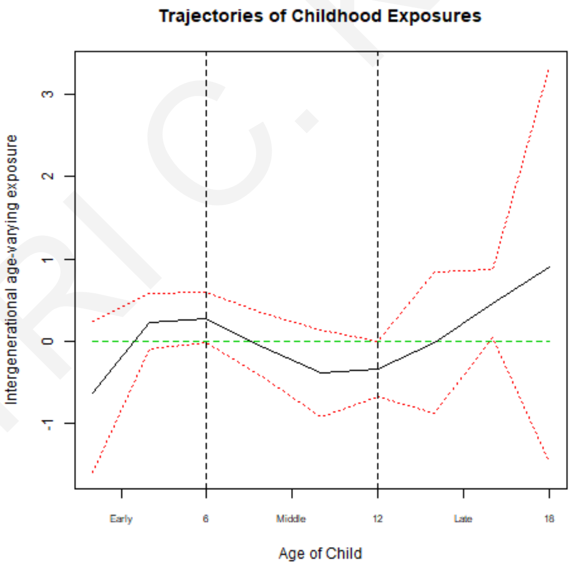
(a) Non-high School Graduates



(b) High School Graduates



(c) Non-college Graduates



(d) College Graduates

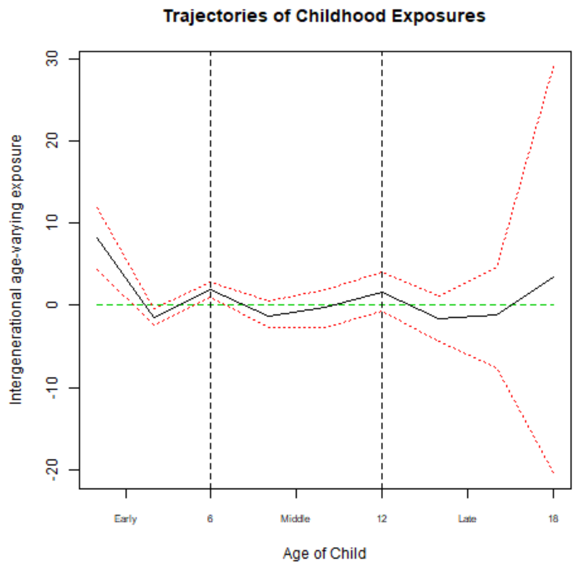
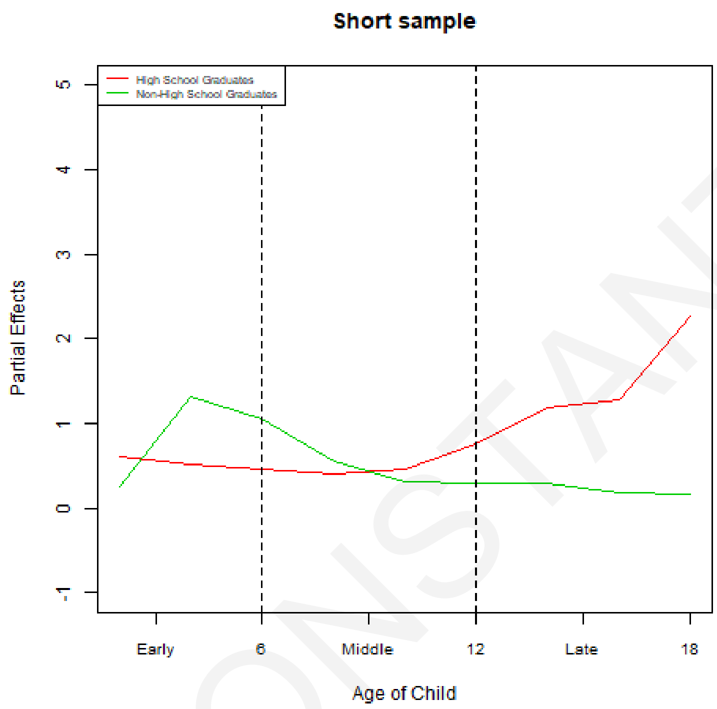


Figure A13: **Partial Effects of Stock of Income by Father's Education**

This figure presents the trajectory partial effects of equation (1.10)based on the short sample for sub-samples based on father's education.

(a) Based on High School Graduation



(b) Based on College Graduation

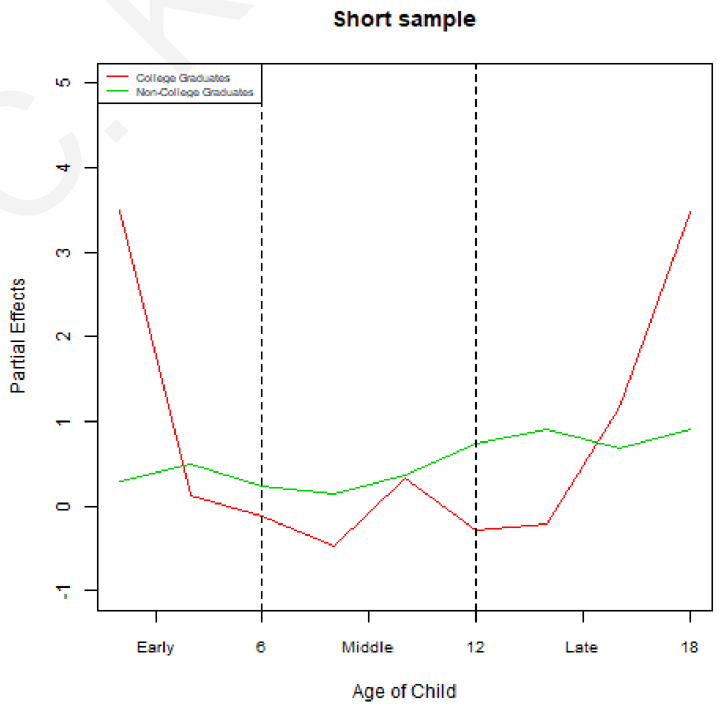
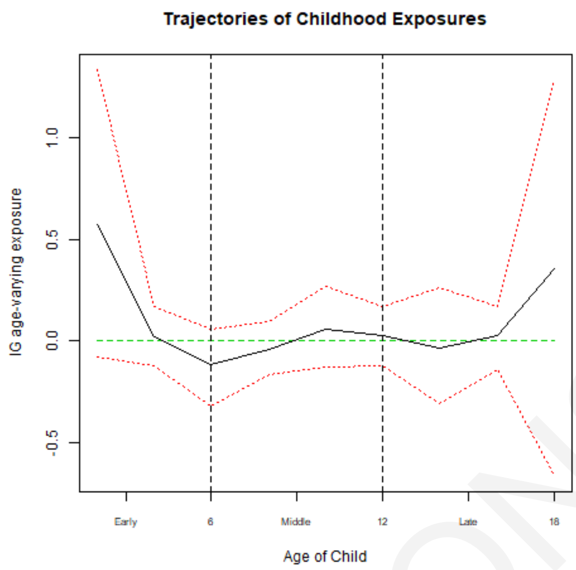


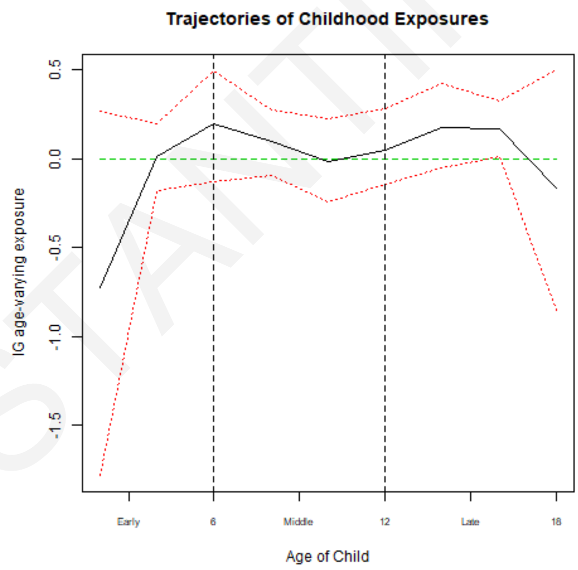
Figure A14: **Intergenerational Trajectories by Family Structure**

This figure presents

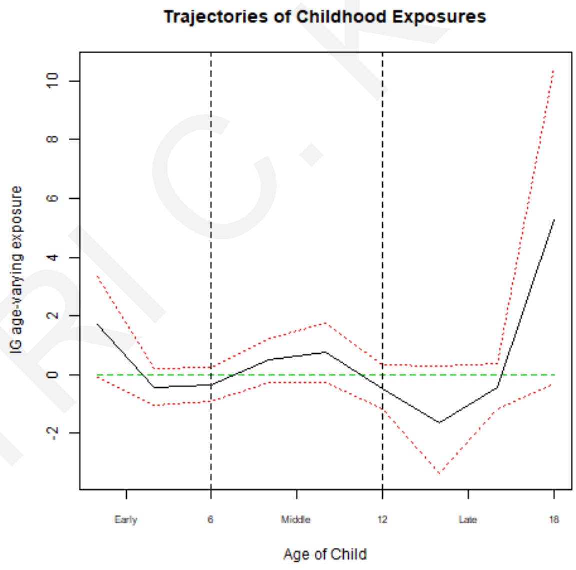
(a) Early Shock - flows



(b) Late Shock - flows



(c) Early Shock - stock



(d) Late Shock - stock

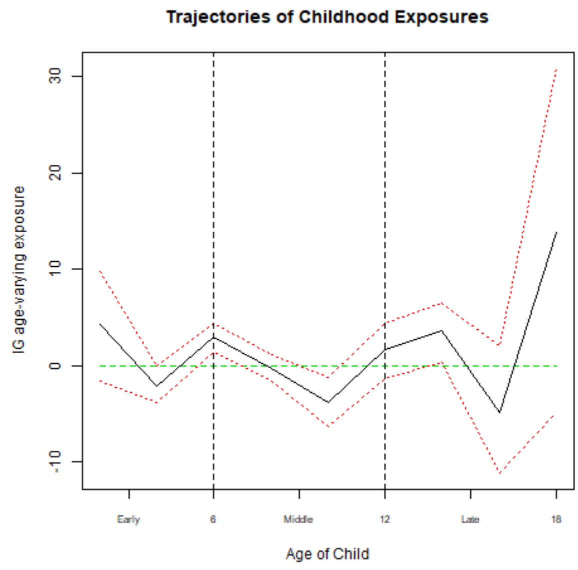
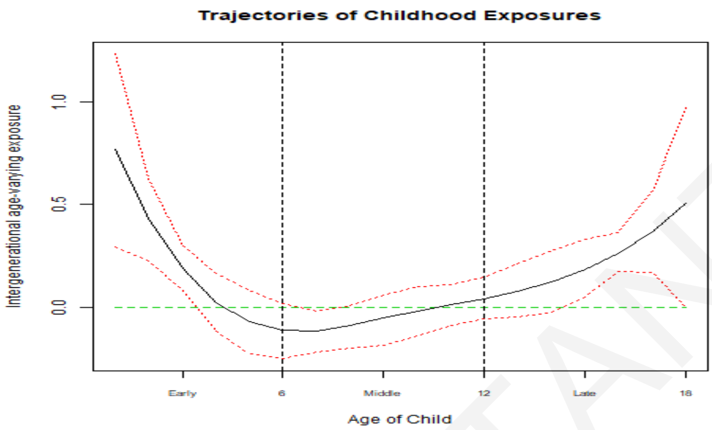
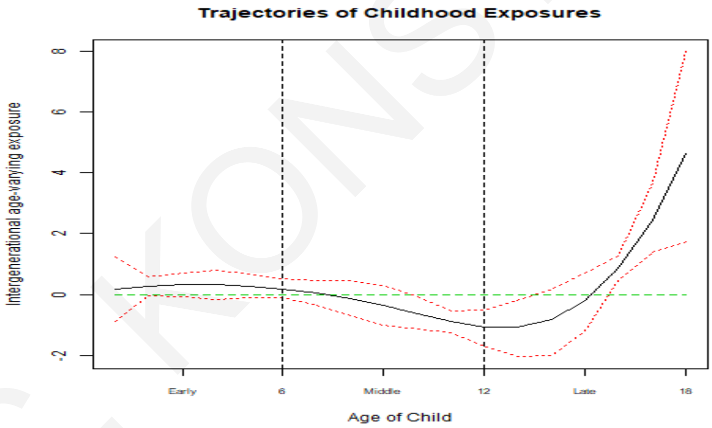


Figure A15: **Schooling Attainment: Intergenerational Trajectories (Short Sample)**

(a) Bi-annual Income



(b) Stock of Income



(c) Partial Effects of Stock of Income

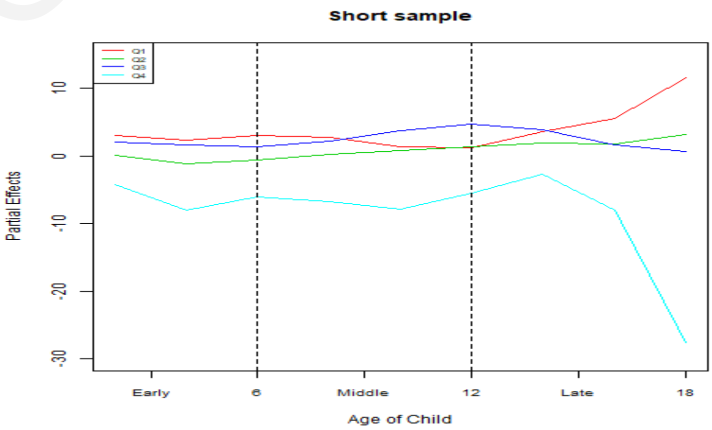
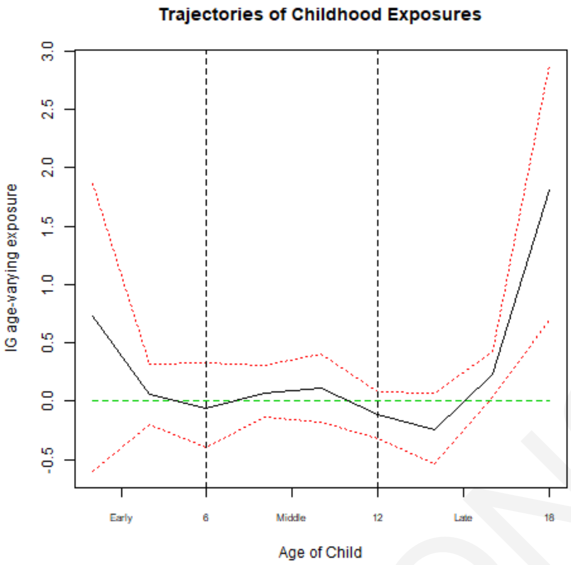
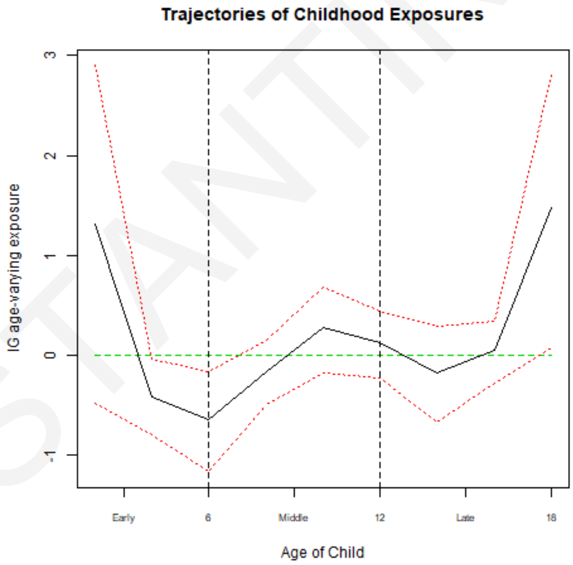


Figure A16: **Schooling Attainment: Intergenerational Trajectories of Income by Parental Income Quartiles - Schooling Attainment (Short Sample)**

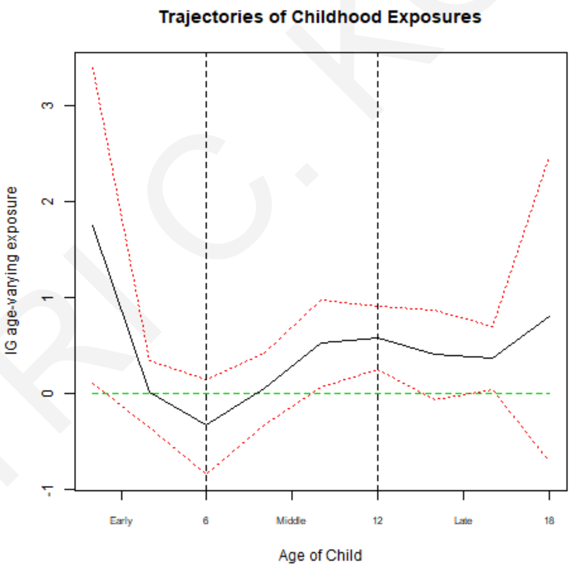
(a) First Quartile



(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

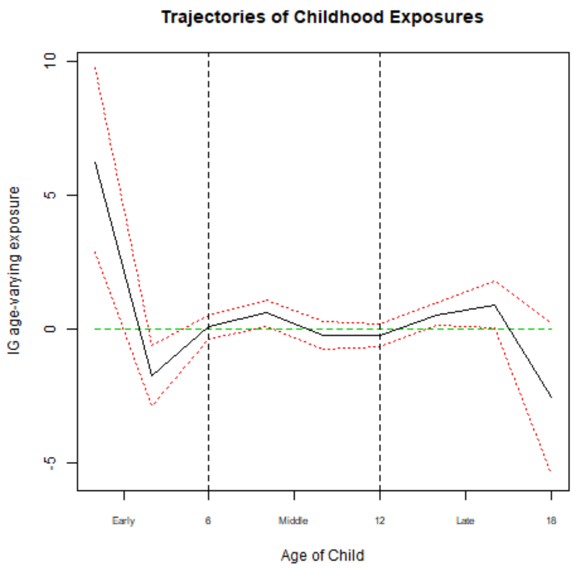
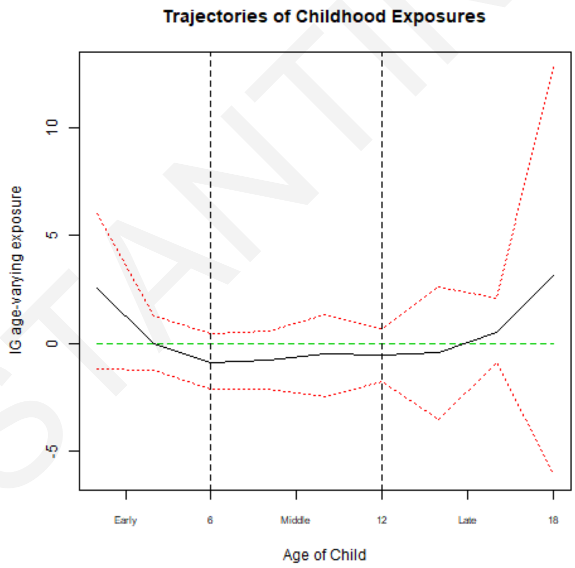


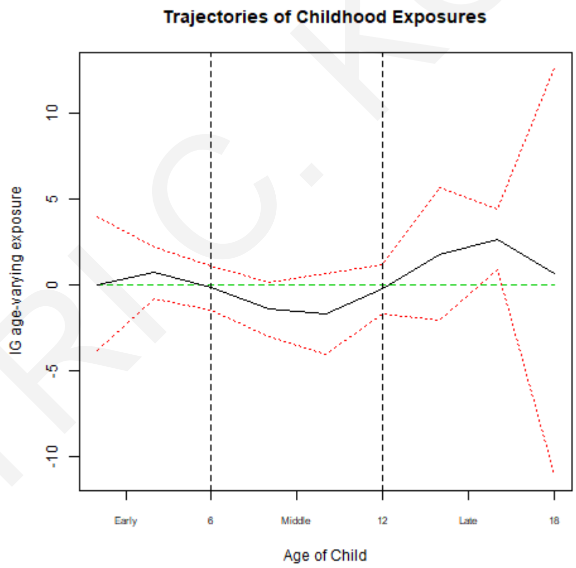
Figure A17: **Schooling Attainment: Intergenerational Trajectories of the Stock of Income by Parental Income Quartiles (Short Sample)**

(a) First Quartile

(b) Second Quartile



(c) Third Quartile



(d) Fourth Quartile

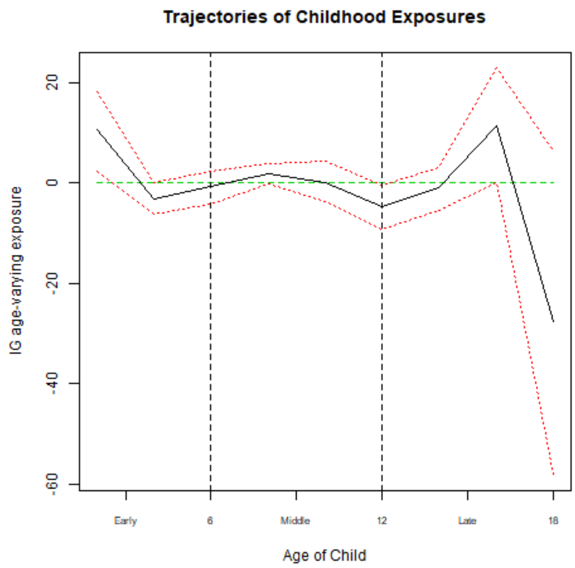


Figure A18: **Schooling Attainment: Partial Effects of the Stock of Income based by Parental Income Quartiles (Short Sample)**

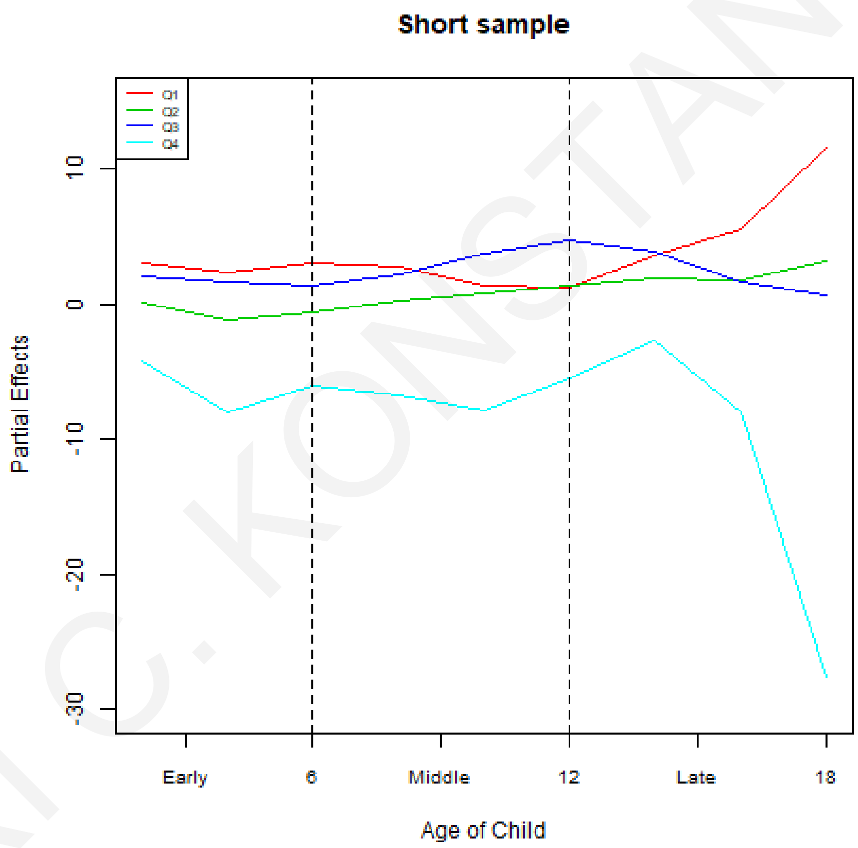
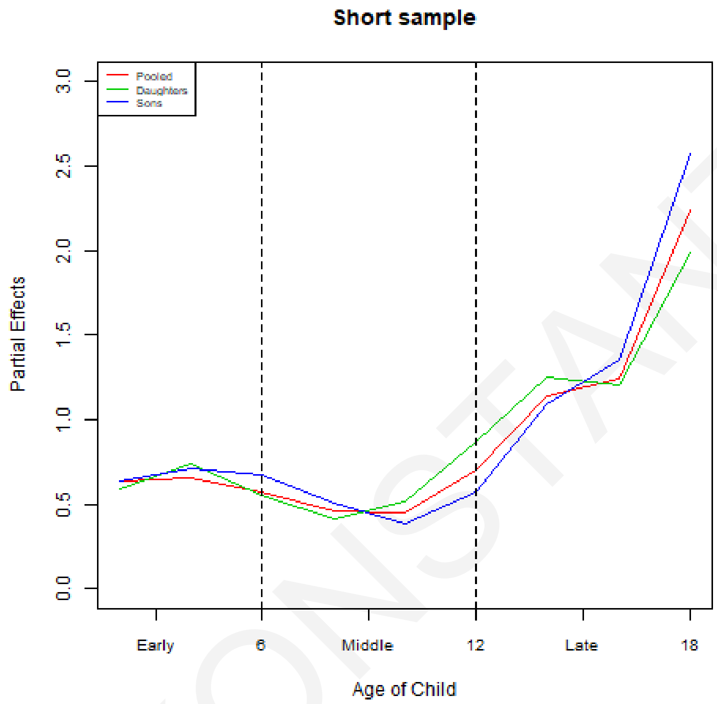


Figure A19: **Partial Effects of the Stock of Income**

This figure presents the trajectories of partial effects of stock of income in equation (1.10) for the short and long samples. The red line corresponds to the baseline (pooled) sample, the green line to the daughters and the blue line to sons'.

(a) Short Sample



(b) Long Sample

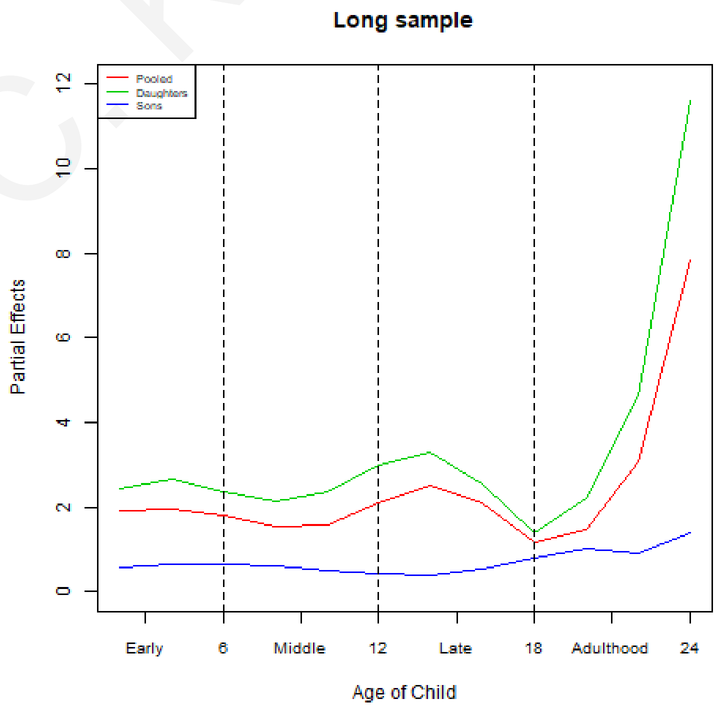


Figure A20: **Intergenerational Trajectories of Growth rates (Short Sample)**

This figure presents the trajectory of income growth experiences.

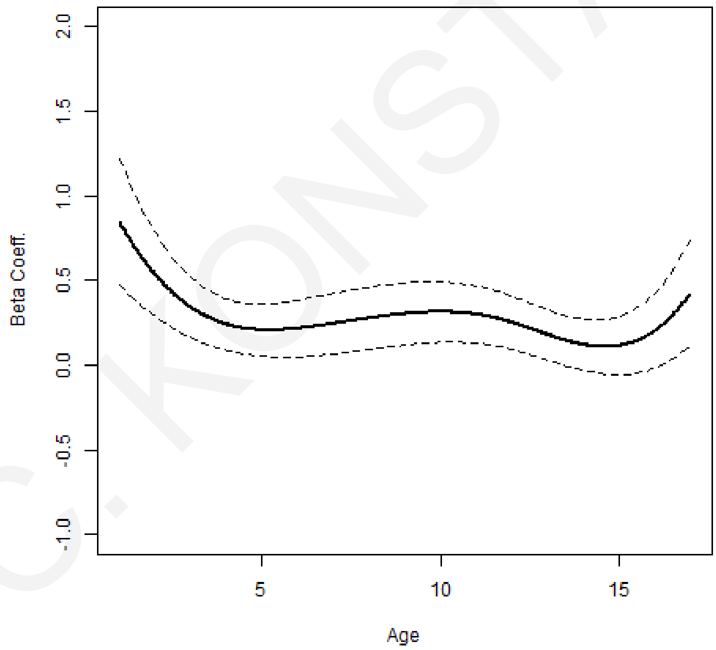
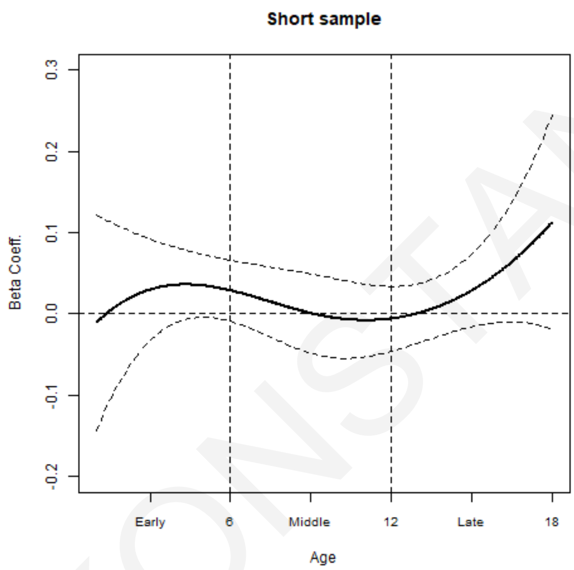


Figure A21: **Intergenerational Trajectories of Stock of Income by Father's Education (Short Sample)**

This figure presents the baseline results from model (1.5) and (1.10) that include parent's education for annual income and stock of income and for both short and long samples.

(a) Annual Income



(c) Stock of Income

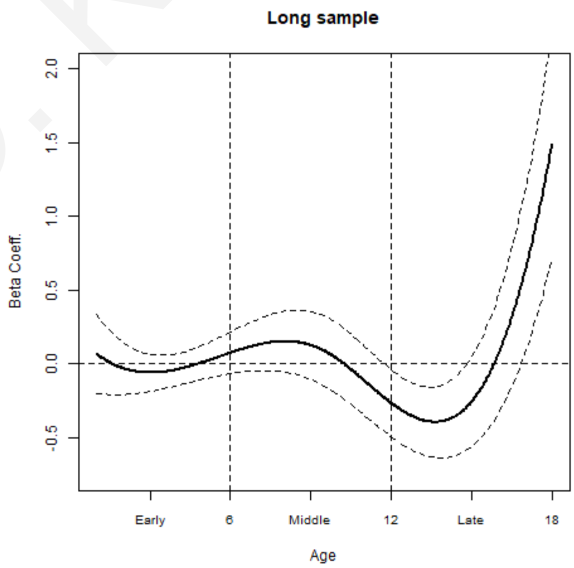


Figure A22: **Schooling Attainment: Intergenerational Trajectories of the Stock of Income by Parental Income Quartiles (Short Sample)**

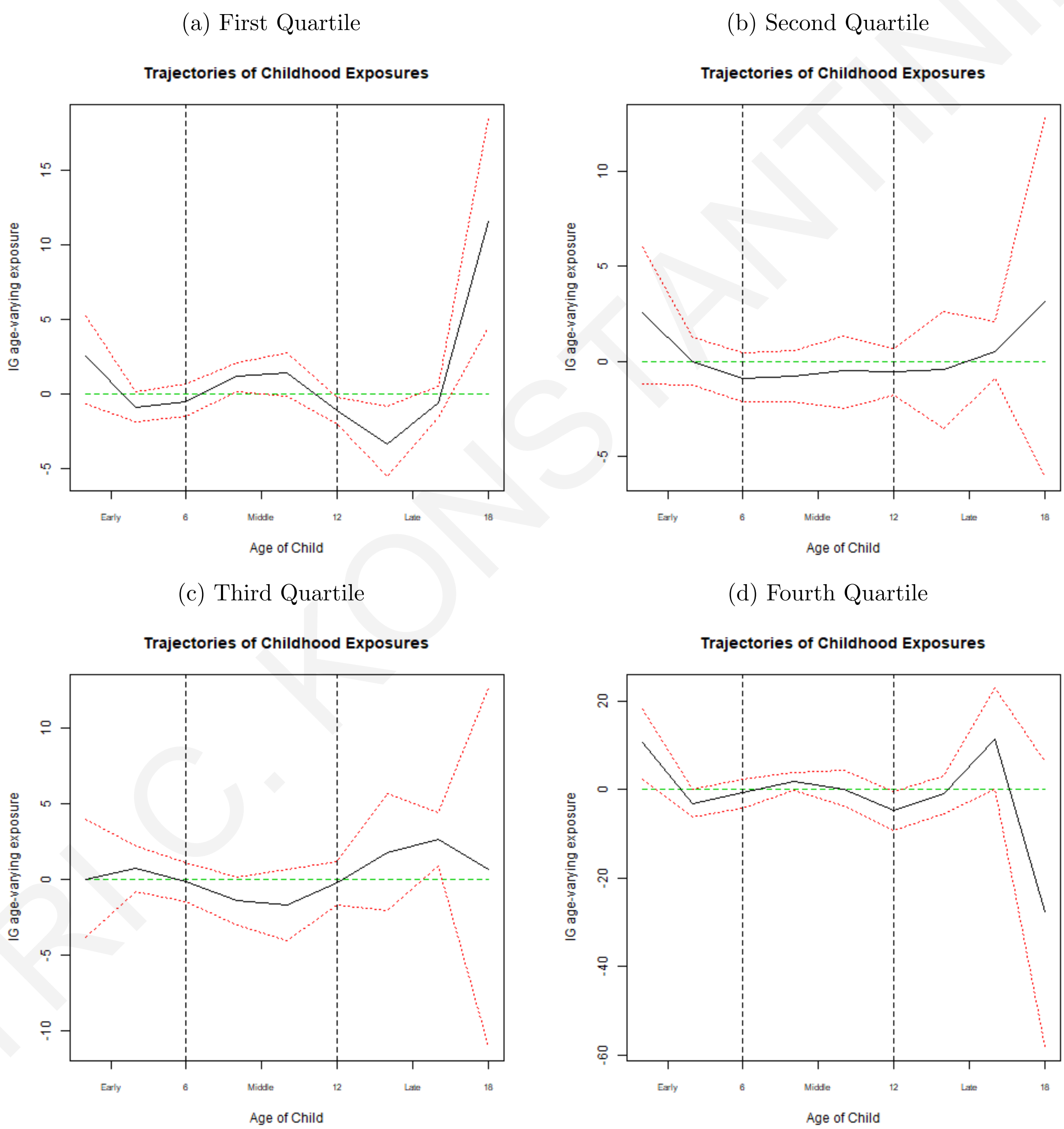


Figure A23: **Schooling Attainment: Partial Effects of the Stock of Income based by Parental Income Quartiles (Short Sample)**

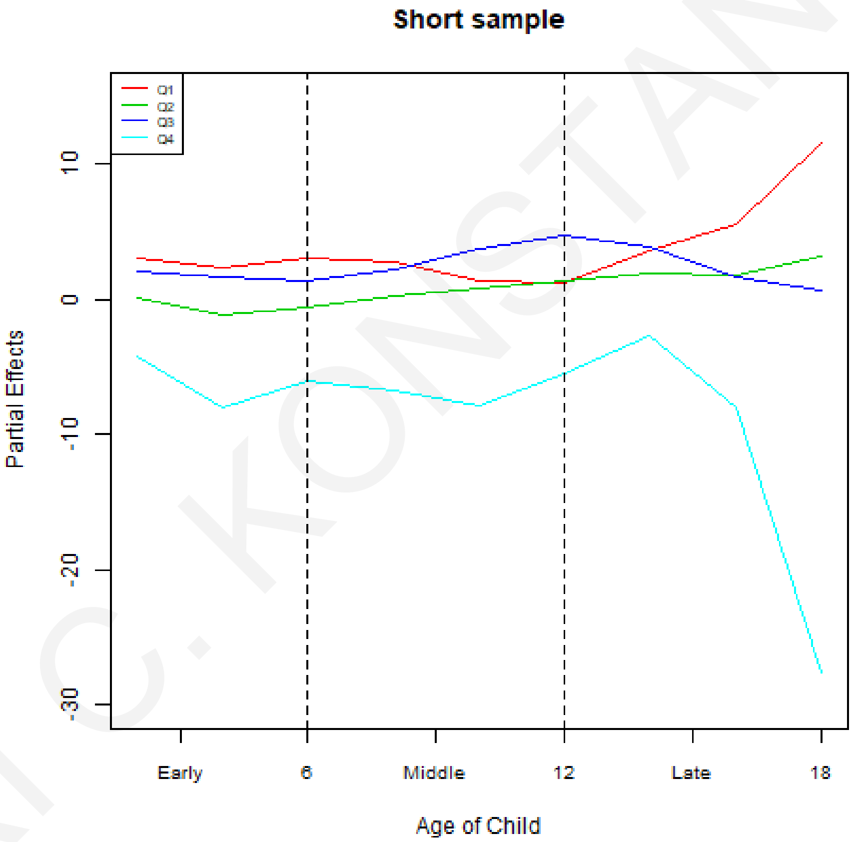
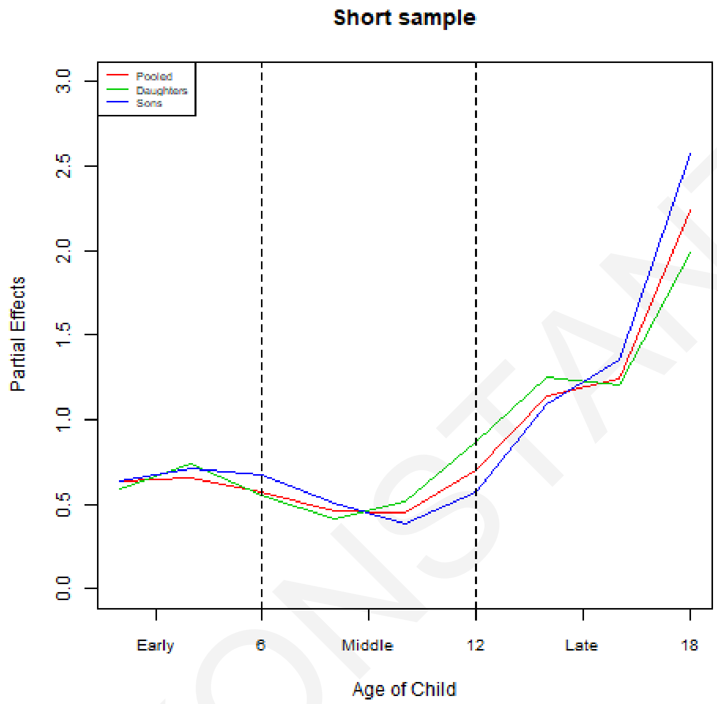


Figure A24: **Partial Effects of the Stock of Income**

This figure presents the trajectories of partial effects of stock of income in equation (1.10) for the short and long samples. The red line corresponds to the baseline (pooled) sample, the green line to the daughters and the blue line to sons'.

(a) Short Sample



(b) Long Sample

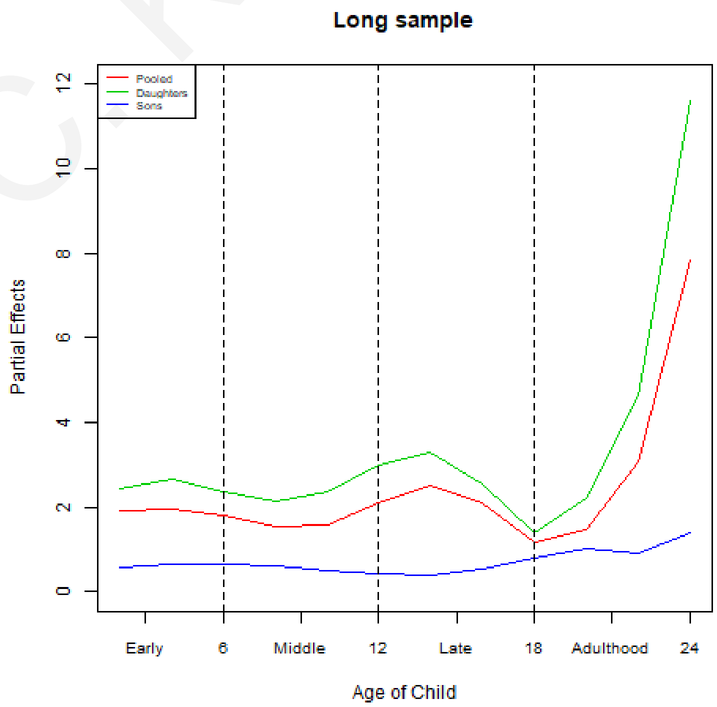


Figure A25: **Intergenerational Trajectories of Growth rates (Short Sample)**

This figure presents the trajectory of income growth experiences.

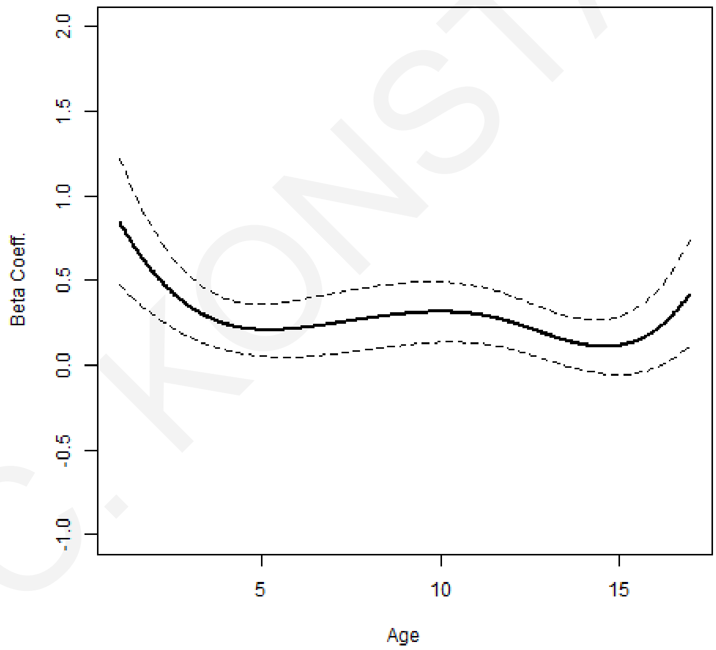
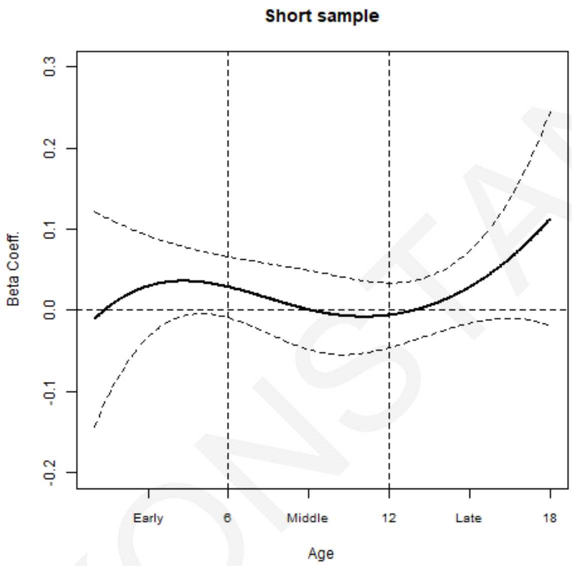


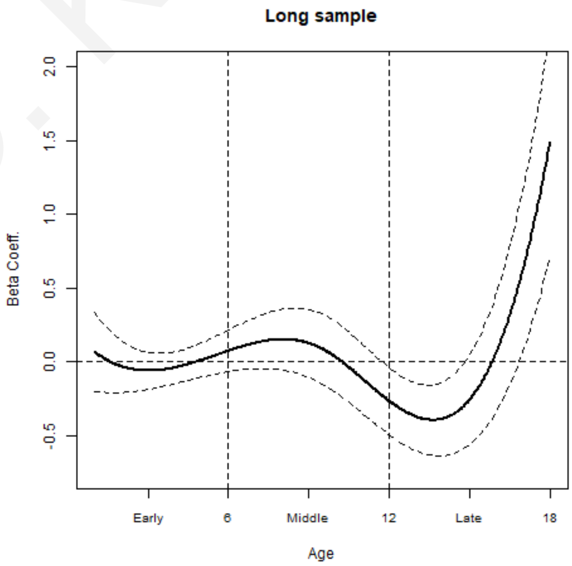
Figure A26: **Intergenerational Trajectories of Stock of Income by Father's Education (Short Sample)**

This figure presents the baseline results from model (1.5) and (1.10) that include parent's education for annual income and stock of income and for both short and long samples.

(a) Annual Income



(c) Stock of Income



Appendix B

B6.1 Derivatives

Lemma B1 *Suppose that A_n and B_n are $n \times k_A$ and $n \times k_B$ matrices with column sums uniformly bounded in absolute value. Then*

1. $\frac{\partial E(A_n' W_n(\lambda) B_n)}{\partial \lambda} = A_n' W_n B_n f(\lambda)$
2. $\frac{\partial E(A_n' W_{n,\lambda}^{-} B_n)}{\partial \lambda} = A_n' W_n B_n f(\lambda)$
3. $\frac{\partial E(A_n' W_n^{-}(\lambda) B_n)}{\partial \lambda} = 2F(\lambda) f(\lambda) A_n' W_n B_n$

Proof: Note that under Assumption 2 we have

$$\begin{aligned} \frac{dE\left[1\left(z_{j,n} \leq \lambda\right)\right]}{d\lambda} &= f(\lambda) \\ \frac{dE\left[1\left(z_{i,n} \leq \lambda\right)1\left(z_{j,n} \leq \lambda\right)\right]}{d\lambda} &= \frac{d\left[F(\lambda)\right]^2}{d\lambda} = 2F(\lambda) f(\lambda) \end{aligned}$$

for any $i \neq j$.

Then, applying Property 17.1(a) in Seber (2008) gives

$$\frac{\partial E(A_n' W_n(\lambda) B_n)}{\partial \lambda} = A_n' \frac{\partial E(W_n(\lambda))}{\partial \lambda} B_n = f(\lambda) A_n' W_n B_n$$

$$\begin{aligned} \frac{\partial E(A_n' W_{n,\lambda}^{-} B_n)}{\partial \lambda} &= A_n' \frac{\partial\ E(W_{n,\lambda}^{-})}{\partial \lambda} B_n = f(\lambda) A_n' W_n B_n \\ \frac{\partial E(A_n' W_n^{-}(\lambda) B_n)}{\partial \lambda} &= A_n' \frac{\partial\ E(W_n^{-}(\lambda))}{\partial \lambda} B_n = 2F(\lambda) f(\lambda) A_n' W_n B_n \end{aligned}$$

as A_n, B_n and W_n are all nonstochastic and the (i,j)th element of $E(W_n(\lambda))$, $E(W_{n,\lambda}^{-})$, and $E(W_n^{-}(\lambda))$ equal $w_{ij,n}(\lambda) = w_{ij,n} 1\left(z_{j,n} \leq \lambda\right)$, $w_{ij,n}(\lambda) = w_{ij,n} 1\left(z_{j,n} \leq \lambda\right)$, and $w_{ij,n} 1\left(z_{i,n} \leq \lambda\right) 1\left(z_{j,n} \leq \lambda\right)$, respectively. This completes the proof of this lemma.

B6.2 Structural Error

First, we obtain the structural error

$$e_n(\theta) = S_n(\theta_y, \lambda)Y_n - X_n\left(W_n\right)\theta_\beta - X_n\left(W_n, \lambda\right)\delta_{\theta_\beta}, \tag{B1}$$

where

$$\begin{aligned} S_n(\theta_y,\lambda) =& \, S_n + (\alpha_0 - \alpha)W_n + (\delta_{\alpha_{2,0}} - \delta_{\alpha_2})W_n(\lambda_0) + (\delta_{\alpha_{1,0}} - \delta_{\alpha_1})W^-_{n,\lambda_0} \\ &+(\delta_{\alpha\alpha 0} - \delta_{\alpha\alpha})W^-_n(\lambda_0) + \delta_{\alpha_2}(W_n(\lambda_0) - W_n(\lambda)) + \delta_{\alpha_1}(W^-_{n,\lambda_0} - W^-_{n,\lambda}) \\ &+\delta_{\alpha\alpha}(W^-_n(\lambda_0) - W^-_n(\lambda)), \end{aligned} \tag{B2}$$

and from the reduced form model (2.5) we have

$$\begin{aligned} Y_n &= S_n^{-1}\left[X_n(W_n)\theta_{\beta_0} + X_n(W_n\lambda_0)\delta_{\theta_{\beta_0}} + e_n\right] \\ &= X_n(W_n)\theta_{\beta_0} + \left[\alpha_0 W_n + \delta_{\alpha_{2,0}}W_n(\lambda_0) + \delta_{\alpha_{1,0}}W^-_{n,\lambda_0} + \delta_{\alpha\alpha,0}W^-_n(\lambda_0)\right] S_n^{-1}X_n(W_n)\theta_{\beta_0} \\ &+ X_n(W_n\lambda_0)\delta_{\theta_{\beta_0}} + \left[\alpha_0 W_n + \delta_{\alpha_{2,0}}W_n(\lambda_0) + \delta_{\alpha_{1,0}}W^-_{n,\lambda_0} + \delta_{\alpha\alpha,0}W^-_n(\lambda_0)\right] S_n^{-1}X_n(W_n,\lambda_0)\delta_{\theta_{\beta_0}} \\ &+ S_n^{-1}e_n \\ &= X_n(W_n)\theta_{\beta_0} + \left[\alpha_0 G_n + \delta_{\alpha_{2,0}}G_n(\lambda_0) + \delta_{\alpha_{1,0}}G^-_{n,\lambda_0} + \delta_{\alpha\alpha,0}G^-_n(\lambda_0)\right] X_n(W_n)\theta_{\beta_0} \\ &+ X_n(W_n\lambda_0)\delta_{\theta_{\beta_0}} + \left[\alpha_0 G_n + \delta_{\alpha_{2,0}}G_n(\lambda_0) + \delta_{\alpha_{1,0}}G^-_{n,\lambda_0} + \delta_{\alpha\alpha,0}G^-_n(\lambda_0)\right] X_n(W_n\lambda_0)\delta_{\theta_{\beta_0}} \\ &+ S_n^{-1}e_n, \end{aligned}$$

where the second equality follows from $(I_n - A)(I_n - A)^{-1} = I_n$ which implies $(I_n - A)^{-1} = I_n + A(I_n - A)^{-1}$ and the third equality uses the definitions of $G_n = W_n S_n^{-1}$, $G_n(\lambda) = W_n(\lambda)S_n^{-1}$, $G^-_{n,\lambda} = W^-_{n,\lambda}S_n^{-1}$, and $G^-_n(\lambda) = W^-_n(\lambda)S_n^{-1}$.

As defined in Section 2, $X^*_{n,\lambda} = [X_n(W_n), X_n(W_n,\lambda)]$, $X^*_n = [X_n(W_n), X_n(W_n,\lambda_0)]$, $\theta_{y0} = (\alpha_0, \delta_{\alpha_{2,0}}, \delta_{\alpha_{1,0}}, \delta_{\alpha_0\alpha_0})'$ and $\theta^*_0 = (\theta'_{\beta_0}, \delta'_{\theta_{\beta_0}})'$, model (B3) can be rewritten as

$$Y_n = X^*_n\theta^*_0 + \left[G_n X^*_n\theta^*_0, G_n(\lambda_0)X^*_n\theta^*_0, G^-_{n,\lambda_0}X^*_n\theta^*_0, G^-_n(\lambda_0)X^*_n\theta^*_0\right]\theta_{y0} + S_n^{-1}e_n. \tag{B3}$$

Substituting (B2) and (B3) into equation (B1) gives

$$e_n(\theta) = d_n(\theta) + \left[A_n(\theta_y,\lambda) + I_n\right]e_n \tag{B4}$$

where

$$\begin{aligned} d_n(\theta) =& \, X^*_n(\theta^*_0 - \theta^*) + (X^*_n - X^*_{n,\lambda})\theta^* + (\alpha_0 - \alpha)G_n X^*_n\theta^*_0 + (\delta_{\alpha_{2,0}} - \delta_{\alpha_2})G_n(\lambda_0)X^*_n\theta^*_0 \\ &+(\delta_{\alpha_{1,0}} - \delta_{\alpha_1})G^-_{n,\lambda_0}X^*_n\theta^*_0 + (\delta_{\alpha\alpha,0} - \delta_{\alpha\alpha})G^-_n(\lambda_0)X^*_n\theta^*_0 + \delta_{\alpha_2}(G_n(\lambda_0) - G_n(\lambda))X^*_n\theta^*_0 \\ &+\delta_{\alpha_1}(G^-_{n,\lambda_0} - G^-_{n,\lambda})X^*_n\theta^*_0 + \delta_{\alpha\alpha}(G_n(\lambda_0)^- - G^-_n(\lambda))X^*_n\theta^*_0 \end{aligned}$$

and

$$A_n(\theta_y,\lambda)= (\alpha_0-\alpha)G_n + (\delta_{\alpha 2,0}-\delta_{\alpha_2})G_n(\lambda_0) + (\delta_{\alpha 1,0}-\delta_{\alpha_1})G^-_{n,\lambda_0} + (\delta_{\alpha\alpha,0}-\delta_{\alpha\alpha})G^-_n(\lambda_0)$$
$$+\delta_{\alpha_2}(G_n(\lambda_0)-G_n(\lambda)) + \delta_{\alpha_1}(G^-_{n,\lambda_0}-G^-_{n,\lambda}) + \delta_{\alpha\alpha}(G^-_n(\lambda_0)-G^-_n(\lambda)).$$

Next, we obtain the identification condition.

B6.3 Identification Conditions

Substituting (B4) into (2.6) and taking expectation, we obtain the $(m+k_Q)\times 1$ column vector of linear and quadratic moments

$$E(g_n(\theta)) = \begin{bmatrix} E(d_n(\theta)'P_{1n}d_n(\theta)) + tr(\Gamma_n E(A_n(\theta_y,\lambda)P^s_{1n})) + tr(\Gamma_n E(A_n(\theta_y,\lambda)'P_{1n}A_n(\theta_y,\lambda)) \\ . \\ . \\ . \\ E(d_n(\theta)'P_{mn}d_n(\theta)) + tr(\Gamma_n E(A_n(\theta_y,\lambda)P^s_{mn})) + tr(\Gamma_n E(A_n(\theta_y,\lambda)'P_{mn}A_n(\theta_y,\lambda)) \\ E(Q'_n d_n(\theta)) \end{bmatrix}. \quad \text{(B5)}$$

The identification condition states that $a_0 \lim_{n\to\infty} n^{-1}E(g_n(\theta)) = 0$ has a unique root at $\theta_0 \in \Theta$. Applying Taylor expansion gives

$$E\left[g_n(\theta)\right] = E\left[g_n(\theta_0)\right] + \frac{\partial E\left[g_n(\bar{\theta})\right]}{\partial \theta'}(\theta-\theta_0)$$

where $\bar{\theta}$ lies between θ and θ_0.
Evidently, there will be a unique θ_0 satisfying $a_0 \lim_{n\to\infty} n^{-1}E(g_n(\theta)) = 0$ if $\partial E(g_n(\theta))/\,\partial\theta'$ has a full rank k_θ over $\theta\in\Theta$.

Firstly, we calculate $\partial E(Q_n d_n(\theta))/\,\partial\theta'$, where

$$\begin{aligned} E(Q'_n d_n(\theta)) &= E\{Q'_n[X^*_n(\theta^*_0-\theta^*) + (X^*_n - X^*_{n,\lambda})\theta^* + (\alpha_0-\alpha)G_n X^*_n\theta^*_0 \\ &+ (\delta_{\alpha_{2,0}}-\delta_{\alpha_2})G_n(\lambda_0)X^*_n\theta^*_0 + (\delta_{\alpha_{1,0}}-\delta_{\alpha_1})G^-_{n,\lambda_0}X^*_n\theta^*_0 \\ &+ (\delta_{\alpha\alpha 0}-\delta_{\alpha\alpha})G^-_n(\lambda_0)X^*_n\theta^*_0 + \delta_{\alpha_2}(G_n(\lambda_0)-G_n(\lambda))X^*_n\theta^*_0 \\ &+ \delta_{\alpha_1}(G^-_{n,\lambda_0}-G^-_{n,\lambda})X^*_n\theta^*_0 + \delta_{\alpha\alpha}(G_n(\lambda_0)^- - G^-_n(\lambda))X^*_n\theta^*_0]\} \end{aligned}$$

Applying straightforward calculations gives

$$
\begin{aligned}
G_{\theta_y}(\lambda) &= \frac{\partial E(Q_n' d_n(\theta))}{\partial \theta_y'} = -E\left[Q_n' G_n X_n^* \theta_0^*, Q_n' G_n(\lambda) X_n^* \theta_0^*, Q_n' G_{n,\lambda}^- X_n^* \theta_0^*, Q_n' G_n^-(\lambda) X_n^* \theta_0^*\right] \\
G_{\theta_\beta} &= \frac{\partial E(Q_n' d_n(\theta))}{\partial \theta_\beta'} = -E\left[Q_n' X_n(W_n)\right] \\
G_{\delta_{\theta_\beta}}(\lambda) &= \frac{\partial E(Q_n' d_n(\theta))}{\partial \delta_{\theta_\beta}'} = -E\left[Q_n' X_n(W_n, \lambda)\right].
\end{aligned}
$$

Also, we obtain

$$
\begin{aligned}
& G_\lambda(\delta_\alpha, \delta_{\theta_\beta}, \lambda) = \frac{\partial E(Q_n' d_n(\theta))}{\partial \lambda} \\
= \; & -f(\lambda)\left[E(Q_n' X_n|\lambda), E(Q_n' W_n X_n|\lambda), E(Q_n' W_n X_n|\lambda), E(Q_n W_n(\lambda) X_n|\lambda) + E(Q_n W_{n,\lambda}^- X_n|\lambda)\right] \delta_{\theta_\beta} \\
& -f(\lambda)\left[E(Q_n' G_n X_n^* \theta_0^*|\lambda), E(Q_n' G_n X_n^* \theta_0^*|\lambda), E(Q_n G_n(\lambda) X_n^* \theta_0^*|\lambda) + E(Q_n G_{n,\lambda}^- X_n^* \theta_0^*|\lambda))\right] \delta_\alpha.
\end{aligned}
$$

where applying Lemma B1 gives

$$
\begin{aligned}
& \frac{\partial}{\partial \lambda} E\left[Q_n' X_n(\lambda) \delta_\beta\right] \\
= & \sum_{i=1}^{n} \frac{\partial}{\partial \lambda} E\left[Q_{i,n} x_{i,n}' \delta_\beta 1\{z_i \leq \lambda\}\right] \\
= & \sum_{i=1}^{n} \frac{\partial}{\partial \lambda} \int_{-\infty}^{\lambda} E\left[Q_{i,n} x_{i,n}' \delta_\beta | z_{i,n}\right] f(z_{i,n})\, dz_{i,n} \\
= & \sum_{i=1}^{n} E\left(Q_{i,n} x_{i,n}' \delta_{\theta_\beta} | \lambda\right) f(\lambda) = E\left(Q_n' X_n \delta_\beta | \lambda\right) f(\lambda),
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial}{\partial \lambda} E\left[Q_n' W_n(\lambda) X_n \delta_{\gamma_2}\right] \\
= & \sum_{i=1}^{n} \sum_{j \neq i} w_{ij,n} \frac{\partial}{\partial \lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma_2} 1\{z_{j,n} \leq \lambda\}\right] \\
= & \sum_{i=1}^{n} \sum_{j \neq i} w_{ij,n} \frac{\partial}{\partial \lambda} \int_{-\infty}^{\lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma_2} | z_{j,n}\right] f(z_{j,n})\, dz_{j,n} \\
= & \sum_{i=1}^{n} \sum_{j \neq i} w_{ij,n} E\left(Q_{i,n} x_{j,n}' \delta_{\gamma_2} | \lambda\right) f(\lambda) = E\left(Q_n W_n X_n \delta_{\gamma_2} | \lambda\right) f(\lambda),
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial}{\partial \lambda} E\left[Q_n' W_{n,\lambda}^- X_n \delta_\gamma\right] \\
= & \sum_{i=1}^{n} \sum_{j \neq i} w_{ij,n} \frac{\partial}{\partial \lambda} E\left[Q_{i,n} x_{j,n}' \delta_\gamma 1\{z_i \leq \lambda\}\right] \\
= & \sum_{i=1}^{n} \sum_{j \neq i} w_{ij,n} E\left(Q_{i,n} x_{j,n}' \delta_\gamma | \lambda\right) f(\lambda) = E\left(Q_n W_n X_n \delta_\gamma | \lambda\right) f(\lambda),
\end{aligned}
$$

and applying the Leibniz rule

$$\frac{d}{dx}\int_{a(x)}^{b(x)} g(x,y)\,dy = g(x,b(x))\,b'(x) - g(x,a(x))\,a'(x) + \int_{a(x)}^{b(x)} \frac{dg(x,y)}{dx}dy$$

we obtain

$$\begin{aligned}
&\frac{\partial}{\partial \lambda} E\left[Q_n' W_n^-(\lambda) X_n \delta_{\gamma\gamma}\right] \\
=& \sum_{i=1}^{n}\sum_{j\neq i} w_{ij,n} \frac{\partial}{\partial \lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} 1(z_{i,n} \leq \lambda) 1(z_{j,n} \leq \lambda)\right] \\
=& \sum_{i=1}^{n}\sum_{j\neq i} w_{ij,n} \frac{\partial}{\partial \lambda} \int_{-\infty}^{\lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} 1(z_{i,n} \leq \lambda) | z_{j,n}\right] f(z_{j,n})\, dz_{j,n} \\
=& \sum_{i=1}^{n}\sum_{j\neq i} w_{ij,n} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} 1(z_{i,n} \leq \lambda) | \lambda\right] f(\lambda) \\
+& \sum_{i=1}^{n}\sum_{j\neq i} w_{ij,n} \int_{-\infty}^{\lambda} \frac{\partial}{\partial \lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} 1(z_{i,n} \leq \lambda) | z_{j,n}\right] f(z_{j,n})\, dz_{j,n} \\
=& E\left[Q_n W_{n,\lambda}^- X_n \delta_{\gamma\gamma} | \lambda\right] f(\lambda) \\
+& \sum_{i=1}^{n}\sum_{j\neq i} w_{ij,n} \int_{-\infty}^{\lambda} \left\{ \frac{\partial}{\partial \lambda} \int_{-\infty}^{\lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} | z_{i,n}, z_{j,n}\right] f(z_{i,n})\, dz_{i,n} \right\} f(z_{j,n})\, dz_{j,n} \\
=& E\left[Q_n W_{n,\lambda}^- X_n \delta_{\gamma\gamma} | \lambda\right] f(\lambda) + \sum_{i=1}^{n}\sum_{j\neq i} w_{ij,n} \int_{-\infty}^{\lambda} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} | \lambda, z_{j,n}\right] f(\lambda) f(z_{j,n})\, dz_{j,n} \\
=& E\left(Q_n W_{n,\lambda}^- X_n \delta_{\gamma\gamma} | \lambda\right) f(\lambda) + \sum_{i=1}^{n} w_{ij,n} E\left[Q_{i,n} x_{j,n}' \delta_{\gamma\gamma} I(z_{j,n} \leq \lambda) | \lambda\right] f(\lambda) \\
=& E\left(Q_n W_{n,\lambda}^- X_n \delta_{\gamma\gamma} | \lambda\right) f(\lambda) + E\left(Q_n W_n(\lambda) X_n \delta_{\gamma\gamma} | \lambda\right) f(\lambda).
\end{aligned}$$

We proceed to show that if the rank condition of Assumption (4.1) fails, we can identify $(\theta_{\beta_0}', \delta_{\theta_{\beta_0}}')'$, as long as $\left(\theta_{y_0}', \lambda_0\right)'$ is identified from using the quadratic moments. By (2.5), we have

$$\begin{aligned}
Y_n =& S_n^{-1} X_n^* \theta_0^* + u_n = (I_n - S_n) S_n^{-1} X_n^* \theta_0^* + X_n^* \theta_0^* + u_n \\
&= [G_n X_n^* \theta_0^*, G_n(\lambda_0) X_n^* \theta_0^*, G_{n,\lambda_0}^- X_n^* \theta_0^*, G_n^-(\lambda_0) X_n^* \theta_0^*](\alpha_0, \delta_{\alpha_{1,0}}, \delta_{\alpha_{2,0}}, \delta_{\alpha\alpha,0})' + X_n^* \theta_0^* + u_n
\end{aligned} \tag{B6}$$

where $u_n = S_n^{-1} e_n$ and u_n therefore follows a SAR model, $u_n = (\alpha_0 W_n + \delta_{\alpha_{1,0}} W_{n,\lambda_0}^- + \delta_{\alpha_{2,0}} W_n(\lambda_0) + \delta_{\alpha\alpha,0} W_n^-(\lambda_0)) u_n + e_n$. In parallel to Lee (2007) we consider the following example that X_n^* and $[G_n X_n^* \theta_0^*, G_n(\lambda_0) X_n^* \theta_0^*, G_{n,\lambda_0}^- X_n^* \theta_0^*, G_n^-(\lambda_0) X_n^* \theta_0^*]$ are linearly dependent. That is, there exists a $6k_1 \times 4$ non-zero constant matrix c_0 such that

$X_n^* c_0 = [G_n X_n^* \theta_0^*, G_n(\lambda_0) X_n^* \theta_0^*, G_{n,\lambda_0}^- X_n^* \theta_0^*, G_n^-(\lambda_0) X_n^* \theta_0^*]$. Then, model (B6) becomes

$$Y_n = X_n^* \left(c_0 \theta_{y0} + \theta_0^*\right) + u_n, \tag{B7}$$

and

$$\begin{aligned} Q_n' d_n(\theta) &= Q_n' X_n^* \left[\theta_0^* - \theta^* + c_0 \left(\theta_y - \theta_{y0}\right)\right] + Q_n' \left[X_n(W_n, \lambda_0) - X_n(W_n, \lambda)\right] \delta_{\theta_\beta} \\ &+ Q_n' \left[(G_n(\lambda_0) - G_n(\lambda)) X_n^* \theta_0^*, (G_{n,\lambda_0}^- - G_{n,\lambda}^-) X_n^* \theta_0^*, (G_n^-(\lambda_0) - G_n^-(\lambda)) X_n^* \theta_0^*)\right] (\delta_{\alpha_2}, \delta_{\alpha_1}, \delta_{\alpha\alpha}) \end{aligned}$$

and $G_n(Q_n, \theta)$ does not have full rank as

$$\left[G_{\theta_y}(\lambda), G_{\theta_\beta}, G_{\delta_{\theta_\beta}}(\lambda)\right] = -E\left[Q_n' X_n^* c_0, X_n(W_n), Q_n' X_n(W_n, \lambda)\right]$$

are linearly dependent in column. For model (B7), if $\left(\theta_{y0}', \lambda_0\right)'$ is identified from using the quadratic moments, we have $0 = E\left[Q_n' d_n(\theta)\right] = E\left[Q_n' X_n^* \left(\theta_0^* - \theta^*\right)\right]$ so that θ_0^* is identified as X_n^* has full rank.

Secondly, for $j = 1, \ldots, m$, we calculate

$$\frac{\partial E\left[e_n(\theta)' P_{jn} e_n(\theta)\right]}{\partial \theta'} = E\left[e_n(\theta)' P_{jn}^s \frac{\partial\, e_n(\theta)}{\partial \theta'}\right]$$

where we have

$$\frac{\partial\, e_n(\theta)}{\partial \theta'} = \left[\frac{\partial\, e_n(\theta)}{\partial \theta_y'}, \frac{\partial\, e_n(\theta)}{\partial \theta_\beta'}, \frac{\partial\, e_n(\theta)}{\partial \delta_{\theta_\beta}'}, \frac{\partial\, e_n(\theta)}{\partial \lambda}\right]$$

and

$$\begin{aligned} \frac{\partial\, e_n(\theta)}{\partial \theta_y'} &= \frac{\partial d_n(\theta)}{\partial \theta_y'} + \frac{\partial\left[A_n(\theta_y, \lambda) e_n\right]}{\partial \theta_y'} \\ &= -\left[G_n X_n^* \theta_0^*, G_n(\lambda) X_n^* \theta_0^*, G_{n,\lambda}^- X_n^* \theta_0^*, G_n^-(\lambda) X_n^* \theta_0^*\right] \\ &\quad -\left[G_n e_n, G_n(\lambda) e_n, G_{n,\lambda}^- e_n, G_n^-(\lambda) e_n\right] \end{aligned}$$

$$\frac{\partial\, e_n(\theta)}{\partial \theta_\beta'} = \frac{\partial d_n(\theta)}{\partial \theta_\beta'} + \frac{\partial\left[A_n(\theta_y, \lambda) e_n\right]}{\partial \theta_\beta'} = -X_n(W_n)$$

$$\frac{\partial\, e_n(\theta)}{\partial \delta_{\theta_\beta}'} = \frac{\partial d_n(\theta)}{\partial \delta_{\theta_\beta}'} + \frac{\partial\left[A_n(\theta_y, \lambda) e_n\right]}{\partial \delta_{\theta_\beta}'} = -X_n(W_n, \lambda)$$

and

$$\begin{aligned} \frac{\partial\, e_n(\theta)}{\partial \lambda} &= \frac{\partial d_n(\theta)}{\partial \lambda} + \frac{\partial\left[A_n(\theta_y, \lambda) e_n\right]}{\partial \lambda} \\ &= -\frac{\partial X_{n,\lambda}^*}{\partial \lambda} - \delta_{\alpha_2} \frac{\partial G_n(\lambda) X_n^* \theta_0^*}{\partial \lambda} - \delta_{\alpha_1} \frac{\partial G_{n,\lambda}^- X_n^* \theta_0^*}{\partial \lambda} - \delta_{\alpha\alpha} \frac{\partial G_n^-(\lambda) X_n^* \theta_0^*}{\partial \lambda} \\ &\quad -\left(\delta_{\alpha_2} \frac{\partial G_n(\lambda)}{\partial \lambda} + \delta_{\alpha_1} \frac{\partial G_{n,\lambda}^-}{\partial \lambda} + \delta_{\alpha\alpha} \frac{\partial G_n^-(\lambda)}{\partial \lambda}\right) e_n. \end{aligned}$$

Therefore, we obtain

$$\frac{\partial E\left[e_n\left(\theta\right)' P_{jn} e_n(\theta)\right]}{\partial \theta'} = \Delta_1\left(\theta\right) + \Delta_2\left(\theta\right)$$

where we have

$$\begin{aligned}
&\Delta_1\left(\theta\right) = E\left[d_n(\theta)' P_{jn}^s \frac{\partial\ e_n(\theta)}{\partial \theta'}\right] \\
= \quad & E\left[d_n(\theta)' P_{jn}^s \left[\frac{\partial\ e_n(\theta)}{\partial \theta_y'}, \frac{\partial\ e_n(\theta)}{\partial \theta_\beta'}, \frac{\partial\ e_n(\theta)}{\partial \delta_{\theta_\beta}'}, \frac{\partial\ e_n(\theta)}{\partial \lambda}\right]\right] \\
= \quad & -E\left\{d_n(\theta)' P_{jn}^s \left[G_n X_n^* \theta_0^*, G_n(\lambda) X_n^* \theta_0^*, G_{n,\lambda}^- X_n^* \theta_0^*, G_n^-(\lambda) X_n^* \theta_0^*, X_n\left(W_n\right), X_n\left(W_n, \lambda\right),\right.\right. \\
& \left.\left.\frac{\partial X_{n,\lambda}^*}{\partial \lambda} + \delta_{\alpha_2} \frac{\partial G_n(\lambda) X_n^* \theta_0^*}{\partial \lambda} + \delta_{\alpha_1} \frac{\partial G_{n,\lambda}^- X_n^* \theta_0^*}{\partial \lambda} + \delta_{\alpha\alpha} \frac{\partial G_n^-(\lambda) X_n^* \theta_0^*}{\partial \lambda}\right]\right\} \\
\equiv \quad & -E\left\{d_n(\theta)' P_{jn}^s \left[G_n X_n^* \theta_0^*, G_n(\lambda) X_n^* \theta_0^*, G_{n,\lambda}^- X_n^* \theta_0^*, G_n^-(\lambda) X_n^* \theta_0^*, X_n\left(W_n\right),\right.\right. \\
& \left.\left. X_n\left(W_n, \lambda\right), \varphi_1\left(\theta_y, \lambda\right)\right]\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\Delta_2\left(\theta\right) = E\left[e_n'\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s \frac{\partial\ e_n(\theta)}{\partial \theta'}\right] \\
= \quad & -E\left[e_n'\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s \left[G_n X_n^* \theta_0^*, G_n(\lambda) X_n^* \theta_0^*, G_{n,\lambda}^- X_n^* \theta_0^*, G_n^-(\lambda) X_n^* \theta_0^*,\right.\right. \\
& \left.\left. X_n\left(W_n\right), X_n\left(W_n, \lambda\right), 0_n\right]\right] \\
& -E\left[e_n'\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s \left[G_n e_n, G_n(\lambda) e_n, G_{n,\lambda}^- e_n, G_n^-(\lambda) e_n, 0_{n\times(2k_1)}, 0_{n\times(2k_1)}, \frac{\partial\ e_n(\theta)}{\partial \lambda}\right]\right] \\
= \quad & -E\left\{e_n'\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s \left[G_n e_n, G_n(\lambda) e_n, G_{n,\lambda}^- e_n, G_n^-(\lambda) e_n, 0_{n\times(2k_1)}, 0_{n\times(2k_1)},\right.\right. \\
& \left.\left.\left(\delta_{\alpha_2} \frac{\partial G_n(\lambda)}{\partial \lambda} + \delta_{\alpha_1} \frac{\partial G_{n,\lambda}^-}{\partial \lambda} + \delta_{\alpha\alpha} \frac{\partial G_n^-(\lambda)}{\partial \lambda}\right) e_n\right]\right\} \\
\equiv \quad & -\left[\operatorname{tr}\left(\Gamma_n E\left\{\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s G_n\right\}\right), \operatorname{tr}\left(\Gamma_n E\left\{\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s G_n(\lambda)\right\}\right),\right. \\
& \operatorname{tr}\left(\Gamma_n E\left\{\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s G_{n,\lambda}^-\right\}\right), \operatorname{tr}\left(\Gamma_n E\left\{\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s G_n^-(\lambda)\right\}\right), 0_{n\times(2k_1)}, \\
& \left. 0_{n\times(2k_1)}, \operatorname{tr}\left(\Gamma_n E\left\{\left[A_n(\theta_y, \lambda) + I_n\right]' P_{jn}^s \varphi_2\left(\theta_y, \lambda\right)\right\}\right)\right].
\end{aligned}$$

Taking together the results above gives

$$\left(\frac{\partial E[g_n(\theta)]}{\partial \theta'}\right)' = -\begin{bmatrix}
E\left[d_n(\theta)' P_{1n}^s G_n X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_1(\theta_y, \lambda) G_n\right]\right\} & \dots & E\left[d_n(\theta)' P_{mn}^s G_n X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_m(\theta_y, \lambda) G_n\right]\right\} & E\left[Q_n' G_n X_n^* \theta_0^*\right]' \\
E\left[d_n(\theta)' P_{1n}^s G_n(\lambda) X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_1(\theta_y, \lambda) G_n(\lambda)\right]\right\} & \dots & E\left[d_n(\theta)' P_{mn}^s G_n(\lambda) X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_m(\theta_y, \lambda) G_n(\lambda)\right]\right\} & E\left[Q_n' G_n(\lambda) X_n^* \theta_0^*\right]' \\
E\left[d_n(\theta)' P_{1n}^s G_{n,\lambda}^- X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_1(\theta_y, \lambda) G_{n,\lambda}^-\right]\right\} & \dots & E\left[d_n(\theta)' P_{mn}^s G_{n,\lambda}^- X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_m(\theta_y, \lambda) G_{n,\lambda}^-\right]\right\} & E\left[Q_n' G_{n,\lambda}^- X_n^* \theta_0^*\right]' \\
E\left[d_n(\theta)' P_{1n}^s G_{n,\lambda}^- X_n^* \theta_0^*\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_1(\theta_y, \lambda) G_n^-(\lambda)\right]\right\} & \dots & E\left[d_n(\theta)' P_{mn}^s G_{n,\lambda}^- X_n^* \theta_0^*\right] + \operatorname{tr}\left(\Gamma_n E\left\{\Gamma_n E\left[\chi_m(\theta_y, \lambda) G_n^-(\lambda)\right]\right\}\right) & E\left[Q_n' G_n^-(\lambda) X_n^* \theta_0^*\right]' \\
E\left[d_n(\theta)' P_{1n}^s X_n(W_n)\right]' & \dots & E\left[d_n(\theta)' P_{mn}^s X_n(W_n)\right]' & E\left[Q_n' X_n(W_n)\right]' \\
E\left[d_n(\theta)' P_{1n}^s X_n(W_n, \lambda)\right]' & \dots & E\left[d_n(\theta)' P_{mn}^s X_n(W_n, \lambda)\right]' & E\left[Q_n' X_n(W_n, \lambda))\right]' \\
E\left[d_n(\theta)' P_{1n}^s \varphi_1\left(\theta_y, \lambda\right)\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_1(\theta_y, \lambda) \varphi_2\left(\theta_y, \lambda\right)\right]\right\} & \dots & E\left[d_n(\theta)' P_{mn}^s \varphi_1\left(\theta_y, \lambda\right)\right] + \operatorname{tr}\left\{\Gamma_n E\left[\chi_m(\theta_y, \lambda) \varphi_2\left(\theta_y, \lambda\right)\right]\right\} & G_\lambda(\delta_\alpha, \delta_{\theta_\beta}, \lambda)'
\end{bmatrix} \qquad \text{(B8)}$$

where we denote $\chi_j(\theta_y, \lambda) = [A_n(\theta_y, \lambda) + I_n]' P^s_{jn}$ for $j = 1, \ldots, m$, to make the notation short.

From (B8), we see that $\partial E[g_n(\theta)]/\partial\theta'$ can still be a full rank matrix even if Assumption (4.1) fails to hold. However, the mathematical expression of the global identification condition can be messy. We therefore include the local identification condition in the main context, i.e., Assumption (4.2), which is the condition under which $\partial E[g_n(\theta_0)]/\partial\theta'$ is a full rank matrix. Specifically, we have

$$\left(\frac{\partial E[g_n(\theta_0)]}{\partial\theta'}\right)' = -\begin{bmatrix} \mathrm{tr}\left[\Gamma_n E\left(P^s_{1n}G_n\right)\right] & \ldots & \mathrm{tr}\left[\Gamma_n E\left(P^s_{mn}G_n\right)\right] & E\left[Q_n' G_n X_n^*\theta_0^*\right]' \\ \mathrm{tr}\left\{\Gamma_n E\left[P^s_{1n}G_n(\lambda_0)\right]\right\} & \ldots & \mathrm{tr}\left\{\Gamma_n E\left[P^s_{mn}G_n(\lambda_0)\right]\right\} & E\left[Q_n' G_n(\lambda_0) X_n^*\theta_0^*\right]' \\ \mathrm{tr}\left\{\Gamma_n E\left[P^s_{1n}G^-_{n,\lambda_0}\right]\right\} & \ldots & \mathrm{tr}\left\{\Gamma_n E\left[P^s_{mn}G^-_{n,\lambda_0}\right]\right\} & E\left[Q_n' G^-_{n,\lambda_0} X_n^*\theta_0^*\right]' \\ \mathrm{tr}\left\{\Gamma_n E\left[P^s_{1n}G^-_n(\lambda_0)\right]\right\} & \ldots & \mathrm{tr}\left(\Gamma_n E\left\{\Gamma_n E\left[P^s_{mn}G^-_n(\lambda_0)\right]\right\}\right) & E\left[Q_n' G^-_n(\lambda_0) X_n^*\theta_0^*\right]' \\ 0_{2k_1} & \ldots & 0_{2k_1} & E\left[Q_n' X_n(W_n)\right]' \\ 0_{4k_1} & \ldots & 0_{4k_1} & E\left[Q_n' X_n(W_n, \lambda_0))\right]' \\ \mathrm{tr}\left\{\Gamma_n E\left[P^s_{1n}\varphi_2\left(\theta_{y,0}, \lambda_0\right)\right]\right\} & \ldots & \mathrm{tr}\left\{\Gamma_n E\left[P^s_{mn}\varphi_2\left(\theta_{y,0}, \lambda_0\right)\right]\right\} & G_\lambda(\delta_{\alpha_0}, \delta_{\theta_{\beta_0}}, \lambda_0)' \end{bmatrix} \quad \text{(B9)}$$

since $d_n(\theta_0) = 0$, $A_n(\theta_{y,0}, \lambda_0) = 0_{n\times n}$, and for $j = 1, \ldots, m$, $\chi_j(\theta_{y,0}, \lambda_0) = P^s_{jn}$ and

$$\begin{aligned} & E\left[P^s_{jn}\varphi_2\left(\theta_{y,0}, \lambda_0\right)\right] \\ = \;& \delta_{\alpha_2} E\left[P^s_{jn}\frac{\partial G_n(\lambda_0)}{\partial\lambda}\right] + \delta_{\alpha_1} E\left[P^s_{jn}\frac{\partial G^-_{n,\lambda_0}}{\partial\lambda}\right] + \delta_{\alpha\alpha} E\left[P^s_{jn}\frac{\partial G^-_n(\lambda_0)}{\partial\lambda}\right] \\ = \;& \left(\delta_{\alpha_2} + \delta_{\alpha_1}\right) E\left(P^s_{jn}G_n|\lambda_0\right) f\left(\lambda_0\right) + \delta_{\alpha\alpha} f\left(\lambda_0\right)\left[E\left(P^s_{jn}G^-_{n,\lambda_0}|\lambda_0\right) + E\left(P^s_{jn}G_n\left(\lambda_0\right)|\lambda_0\right)\right]. \end{aligned} \quad \text{(B10)}$$

B6.4 Proofs of Consistency and Asymptotic Normality

Proof of Proposition 2.1: Firstly, we have $E\left[g_n\left(\theta_0\right)\right] = 0$ and the proof given in Appendix B6.3 shows that $E\left[g_n\left(\theta\right)\right] = 0$ if and only $\theta = \theta_0$ under Assumptions 1-4. Hence, applying Lemma 2.3 in Newey and McFadden (1994) implies that $J_n(\theta)$ has a unique minimum value of zero at true parameter value θ_0.

Secondly, we need to show that $\max_{\theta\in\Theta}\left\|n^{-1}a_n g_n(\theta) - n^{-1}a_n E\left[g_n(\theta)\right]\right\| = o_p(1)$. By (B4), we have

$$e_n\left(\theta\right)'\left(\sum_{j=1}^m a_{nj}P_{jn}\right)e_n\left(\theta\right) = d_n'\left(\theta\right)\left(\sum_{j=1}^m a_{nj}P_{jn}\right)d_n\left(\theta\right) + l_n\left(\theta\right) + m_n\left(\theta\right), \quad \text{(B11)}$$

where $l_n\left(\theta\right) = d_n\left(\theta\right)'\left(\sum_{j=1}^m a_{nj}P^s_{jn}\right)\left[A_n(\theta_y, \lambda) + I_n\right]e_n$

and $m_n(\theta) = e_n'[A_n(\theta_y, \lambda) + I_n]' \left(\sum_{j=1}^{m} a_{nj} P_{jn}\right) [A_n(\theta_y, \lambda) + I_n] \, e_n$.

Because W_n, $A_n(\theta_y, \lambda)$ and $\sum_{j=1}^{m} a_{nj} P_{jn}^{s}$ both have finite column sum norm for any $\theta \in \Theta$ and the elements of X_n are uniformly bounded and strictly exogenous, applying Lemma A.4 and Lemma A.3 in Lin and Lee (2010), we obtain $n^{-1} l_n(\theta) = O_p\left(n^{-1/2}\right)$ and $n^{-1} m_n(\theta) = n^{-1} E\left[m_n(\theta)\right] + O_p\left(n^{-1/2}\right)$ uniformly over $\theta \in \Theta$, respectively. Similarly, we have $n^{-1} Q_n' e_n(\theta) = n^{-1} Q_n' d_n(\theta) + O_p\left(n^{-1/2}\right)$ uniformly over $\theta \in \Theta$ by Lemma A.4 in Lin and Lee (2010). As $d_n(\theta)$ is continuous in θ^{**} and λ enters into $d_n(\theta)$ in the form indicator function, the stochastic equicontinueity result is expected to hold for $n^{-1} a_n g_n(\theta) - n^{-1} a_n E\left[g_n(\theta)\right]$. Then, applying Lemma 2.8 in Newey and McFadden (1994), we obtain $\max_{\theta \in \Theta} \left\| n^{-1} a_n g_n(\theta) - n^{-1} a_n E\left[g_n(\theta)\right] \right\| = o_p(1)$ under Assumptions 1-5.

Taking together all the results above implies $\hat{\theta} \xrightarrow{p} \theta_0$ by Theorem 2.1 in Newey and McFadden (1994). This completes the proof of this Proposition.

Proof of Theorem 1: Under Assumptions 1-5, we have shown that $E\left[g_n(\theta)\right] = 0$ is differentiable function of θ and has a unique solution at an interior point $\theta_0 \in \Theta$, and that $\lim_{n\to\infty} H_n \Lambda_n' a_n' a_n \Lambda_n H_n$ is non-singular, where $\Lambda_n = -\partial E(g_n(\theta_0))/\partial \theta'$ and $H_n = diag(I_{6k_1+4}, n^{a})$. Since $\partial E\left[g_n(\theta_0)\right]/\partial \lambda$ linearly depends on $\delta_{\alpha,0}$ and $\delta_{\theta_{\beta_0}}$ by (B9) and the two parameter vectors are of order n^{-a} for some $0 \leq a < \frac{1}{2}$, we use the weight matrix H_n to rescale Λ_n such that $\Lambda_n H_n$ has full column rank. In addition, applying Lemma A.5 in Lin and Lee (2010) and Cramer-Wold theorem, we obtain

$$\frac{1}{\sqrt{n}} a_n g_n(\theta_0) \xrightarrow{d} N(0, \lim_{n\to\infty} \frac{1}{n} a_n \Omega_n a_n'), \tag{B12}$$

where $\Omega_n = Var(g_n(\theta_0))$. Moreover, because $g_n(\theta)$ and $E\left[g_n(\theta)\right]$ are both continuous in θ^{**} and the elements involving the indicator function of λ satisfies Hölder inequality, the empirical process $\sqrt{n}\left[g_n(\theta) - E\left(g_n(\theta)\right)\right]$ is stochastically equicontinuous, which implies

$$\sup_{\|\theta-\theta_0\| \leq h_n} \frac{\sqrt{n} \left\| g_n(\theta) - g_n(\theta_0) - E\left(g_n(\theta)\right) \right\|}{1 + \sqrt{n} \left| \, \|\theta - \theta_0\| \right.} = o_p(1) \tag{B13}$$

for any $h_n \to 0$ as $n \to \infty$. Hence, applying the central limit theorem of Newey and McFadden (1994) (Theorem 7.2) completes the proof of this theorem.

Before we proceed with the proof of Proposition 2.2 we first introduce several different matrix norms for an $n \times m$ matrix $A_n = (a_{ij})$. (i) $\|A_n\|_{\max} = \max_{1\leq i\leq n, 1\leq j\leq n} |a_{ij}|$. (ii) $\|A_n\|_{sp}$ denotes the spectral norm of A_n. (iii) $\|A_n\|_1 = \max_{1\leq j\leq m} \sum_{i=1}^{n} |a_{ij}|$ and $\|A_n\|_{\infty} = \max_{1\leq i\leq n} \sum_{j=1}^{m} |a_{ij}|$. Also, we denote $\rho(B_n) = \max_{1\leq i\leq n} |\lambda_i(B_n)|$ for any square matrix B_n of size n.

Lemma B2 *Under Assumptions 1-5, we have* $\left\|S_n^{-1}\left(\hat{\theta}_y, \hat{\lambda}\right) - S_n^{-1}\right\|_{sp} = o_p(1)$, $\left\|\hat{Q}_n - Q_n\right\|_{sp} = o_p(1)$ *and* $\left\|\hat{P}_{jn} - P_{jn}\right\|_{sp} = o_p(1)$ *for* $j = 1, \ldots, 4$, *where replacing* $\hat{\theta}$ *by* θ_0 *in* $\hat{Q}_n$ *and* $\hat{P}_{jn}$ *gives* Q_n *and* P_{jn}, *respectively.*

Proof: By Theorem 1, $\sqrt{n}H_n^{-1}\left(\hat{\theta} - \theta_0\right) = O_p(1)$, and we estimate $S_n = S_n(\theta_{y,0}, \lambda_0) = I_n - \alpha_0 W_n - \delta_{\alpha_{1,0}} W_{n,\lambda_0}^- - \delta_{\alpha_{2,0}} W_n(\lambda_0) - \delta_{\alpha\alpha,0} W_n^-(\lambda_0)$ by $\hat{S}_n = S_n(\hat{\theta}_y, \hat{\lambda}) = I_n - \hat{\alpha} W_n - \hat{\delta}_{\alpha_1} W_{n,\hat{\lambda}}^- - \hat{\delta}_{\alpha_2} W_n(\hat{\lambda}) - \hat{\delta}_{\alpha\alpha} W_n^-(\hat{\lambda})$. By Property 4.67 in Seber (2008), we have

$$||\hat{S}_n^{-1} - S_n^{-1}||_{sp} = ||\hat{S}_n^{-1}\left(\hat{S}_n - S_n\right)S_n^{-1}||_{sp} \leq ||\hat{S}_n^{-1}||_{sp}||\hat{S}_n - S_n||_{sp}||S_n^{-1}||_{sp}$$

where $||S_n^{-1}||_{sp} \leq \sqrt{||S_n^{-1}||_1||S_n^{-1}||_\infty} \leq M < \infty$ under Assumption (1.1) and

$$||\hat{S}_n^{-1}||_{sp}^2 = \lambda_{\min}^{-1}\left(\hat{S}_n\hat{S}_n'\right) = \lambda_{\min}^{-1}\left(S_n S_n'\right) + O\left(\left\|\hat{S}_n\hat{S}_n' - S_n S_n'\right\|_{sp}\right)$$

using the result given in the footnote.[4] Hence, we have $\|S_n^{-1} - S_n^{-1}\|_{sp} = O_p\left(||\hat{S}_n - S_n||_{sp}\right)$ if $||\hat{S}_n - S_n||_{sp} = o_p(1)$, where

$$\begin{aligned}
& ||\hat{S}_n - S_n||_{sp} \\
\leq\; & |\hat{\alpha} - \alpha_0|\,\|W_n\|_{sp} + \left\|\hat{\delta}_{\alpha_1} W_{n,\hat{\lambda}}^- - \delta_{\alpha_{1,0}} W_{n,\lambda_0}^-\right\|_{sp} \\
& + \left\|\hat{\delta}_{\alpha_2} W_n(\hat{\lambda}) - \delta_{\alpha_{2,0}} W_n(\lambda_0)\right\|_{sp} + \left\|\hat{\delta}_{\alpha\alpha} W_n^-(\hat{\lambda}) - \delta_{\alpha\alpha,0} W_n^-(\lambda_0)\right\|_{sp} \\
=\; & O_p\left(n^{-1/2}\right) + O_p\left(n^{1/2-a}\right) + \left|\hat{\delta}_{\alpha_1}\right| \left\|W_{n,\hat{\lambda}}^- - W_{n,\lambda_0}^-\right\|_{sp} + \left|\hat{\delta}_{\alpha_2}\right| \left\|W_n(\hat{\lambda}) - W_n(\lambda_0)\right\|_{sp} \\
& + \left|\hat{\delta}_{\alpha\alpha}\right| \left\|W_n^-(\hat{\lambda}) - W_n^-(\lambda_0)\right\|_{sp} \qquad \text{(B14)}
\end{aligned}$$

under Assumption (1.1) and by Theorem 1, and letting $Z\left(\hat{\lambda}, \lambda_0\right)$ be an $n \times n$ diagonal matrix with a typical element equal to $\chi_i\left(\hat{\lambda}, \lambda_0\right) = 1\{z_{i,n} \leq \hat{\lambda}\} - 1\{z_{i,n} \leq \lambda_0\}$, we have

$$\begin{aligned}
W_n^-(\hat{\lambda}) - W_n^-(\lambda_0) &= Z\left(\hat{\lambda}, \lambda_0\right) W_n \\
W_n(\hat{\lambda}) - W_n(\lambda_0) &= W_n Z\left(\hat{\lambda}, \lambda_0\right) \\
W_{n,\hat{\lambda}}^- - W_{n,\lambda_0}^- &= Z\left(\hat{\lambda}, \lambda_0\right) W_n Z\left(\hat{\lambda}, \lambda_0\right).
\end{aligned}$$

[4] For any symmetric matrix A and B of same size, Wely's theorem states that

$$\lambda_{\min}(A) + \lambda_{\min}(B - A) \leq \lambda_{\min}(B) \leq \lambda_{\min}(A) + \lambda_{\max}(B - A)$$

which implies that

$$\lambda_{\min}(B - A) \leq \lambda_{\min}(B) - \lambda_{\min}(A) \leq \lambda_{\max}(B - A)$$

or $|\lambda_{\min}(A) - \lambda_{\min}(B)| \leq \rho(B - A) \leq \|B - A\|_{sp}$ by Property 4.68 (a) in Seber (2008).

By definition we have

$$\left\|Z\left(\hat{\lambda},\lambda_0\right)\right\|_{sp}^2 = \max_{\|\varpi\|=1} \varpi' Z\left(\hat{\lambda},\lambda_0\right)\varpi = \max_{\|\varpi\|=1}\sum_{i=1}^n\sum_{j=1}^n \varpi_i\varpi_j\chi_i\left(\hat{\lambda},\lambda_0\right)\chi_j\left(\hat{\lambda},\lambda_0\right)$$

where $\varpi \neq 0$ is an $n\times 1$ vector. As $\hat{\lambda}-\lambda_0 = O_p\left(n^{a-1/2}\right)$, for any small $\epsilon_n > 0$ there exists a finite constant c_n such that $\Pr\left(\left|\hat{\lambda}-\lambda_0\right| \geq c_n n^{a-1/2}\right) < \epsilon_n$. Letting $B\left(\lambda_0, c_n n^{a-1/2}\right) = \left[\lambda_0 - c_n n^{a-1/2}, \lambda_0 + c_n n^{a-1/2}\right]$, we obtain for some $0<\xi<1/2-a$

$$\begin{aligned}
&\Pr\left\{\max_{\|\varpi\|=1}\sum_{i=1}^n\sum_{j=1}^n \varpi_i\varpi_j\chi_i\left(\hat{\lambda},\lambda_0\right)\chi_j\left(\hat{\lambda},\lambda_0\right) > Mc_n n^{-\xi}\right\}\\
=\;&\Pr\left\{\left[\max_{\|\varpi\|=1}\sum_{i=1}^n\sum_{j=1}^n \varpi_i\varpi_j\chi_i\left(\hat{\lambda},\lambda_0\right)\chi_j\left(\hat{\lambda},\lambda_0\right) > Mc_n n^{-\xi}\right]\cap\left[\hat{\lambda}\in B\left(\lambda_0,c_n n^{a-1/2}\right)\right]\right\}\\
&+\Pr\left\{\left[\max_{\|\varpi\|=1}\sum_{i=1}^n\sum_{j=1}^n \varpi_i\varpi_j\chi_i\left(\hat{\lambda},\lambda_0\right)\chi_j\left(\hat{\lambda},\lambda_0\right) > Mc_n n^{-\xi}\right]\cap\left[\hat{\lambda}\notin B\left(\lambda_0,c_n n^{a-1/2}\right)\right]\right\}\\
\leq\;&\Pr\left\{\sup_{\lambda\in B\left(\lambda_0,c_n n^{a-1/2}\right)}\max_{\|\varpi\|=1}\sum_{i=1}^n\sum_{j=1}^n \varpi_i\varpi_j\chi_i(\lambda,\lambda_0)\chi_j(\lambda,\lambda_0) > Mc_n n^{-\xi}\right\}+\epsilon_n\\
\leq\;&\frac{1}{Mc_n n^{-\xi}}E\left[\max_{\|\varpi\|=1}\sum_{i=1}^n\sum_{j=1}^n \varpi_i\varpi_j\chi_i(\lambda^*,\lambda_0)\chi_j(\lambda^*,\lambda_0)\right]+\epsilon_n\\
=\;&\frac{1}{Mc_n n^{-\xi}}\max_{\|\varpi\|=1}\sum_{i=1}^n \varpi_i^2 E\left[\chi_i^2(\lambda^*,\lambda_0)\right]+\epsilon_n\\
=\;&O\left(n^{a+\xi-1/2}\right)+\epsilon_n = o(1)
\end{aligned}$$

where $\lambda^* = \lambda_0 + c_n n^{a-1/2}$ or $\lambda_0 - c_n n^{a-1/2}$ and it is readily seen that

$$\begin{aligned}
E\left[\chi_i^2(\lambda,\lambda_0)\right] &= F_z(\lambda)+F_z(\lambda_0)-2F_z(\min(\lambda,\lambda_0))\\
&= f\left(\bar{\lambda}\right)(\lambda-\lambda_0)-2f_z\left(\mathring{\lambda}\right)(\min(\lambda,\lambda_0)-\lambda_0)
\end{aligned}$$

with $\bar{\lambda}$ (or $\mathring{\lambda}$) lying between λ (or $\min(\lambda,\lambda_0)$) and λ_0. Hence, we obtain $\left\|Z\left(\hat{\lambda},\lambda_0\right)\right\|_{sp} = O_p\left(n^{-\xi/2}\right)$. Combining the results above gives $||\hat{S}_n - S_n||_{sp} = O_p\left(n^{-1/2}\right)+O_p\left(n^{1/2-a}\right)+O_p\left(n^{-\xi/2}\right) = O_p\left(n^{-\xi/2}\right)$. Then, applying trivial calculations and property of matrix norm will give $\left\|\hat{Q}_n - Q_n\right\|_{sp} = o_p(1)$ and $\left\|\hat{P}_{jn}-P_{jn}\right\|_{sp} = o_p(1)$ for $j=1,\ldots,4$. This completes the proof of this lemma.

Proof of Proposition 2.2: By Lemma A.1 in Lin and Lee (2010), a typical element of $\hat{\Omega}_n$, $\mathrm{tr}\left(\hat{\Gamma}_n\hat{P}_{an}\left(\hat{\Gamma}_n\hat{P}_{bn}\right)^s\right) = \sum_{i=1}^n\sum_{j=1}^n \hat{p}_{an,ij}(\hat{p}_{bn,ij}+\hat{p}_{bn,ji})\sigma_{i,n}^2\sigma_{j,n}^2$, for any $a,b = 1,\ldots,4$, where $\hat{p}_{\varpi n,ij}$ is the (i,j)th element of $\hat{P}_{\varpi n}$ and $\hat{p}_{\varpi n,ii} = 0$ for $\varpi = a,b$ and any i. Denoting $\hat{p}_{\Delta_n,ij} = \hat{p}_{an,ij}(\hat{p}_{bn,ij}+\hat{p}_{bn,ji})$ and $p_{\Delta_n,ij} = p_{an,ij}(p_{bn,ij}+p_{bn,ji})$, we first obtain

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{p}_{\Delta_n,ij}e_{i,n}^2e_{j,n}^2-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}p_{\Delta_n,ij}\sigma_{i,n}^2\sigma_{j,n}^2\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij}-p_{\Delta_n,ij}\right)e_{i,n}^2e_{j,n}^2+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}p_{\Delta_n,ij}\left(e_{i,n}^2e_{j,n}^2-\sigma_{i,n}^2\sigma_{j,n}^2\right)=o_p(1)
\end{aligned} \tag{B15}
$$

where the first term in eq. (B15) equals $o_p(1)$ by using $\|P_{\varpi_n}\|_1 \leq M$, $\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij} = \left(\hat{p}_{an,ij} - p_{an,ij}\right)\left(\hat{p}_{bn,ij} + \hat{p}_{bn,ji} - p_{bn,ij} - p_{bn,ji}\right) + p_{an,ij}\left(\hat{p}_{bn,ij} + \hat{p}_{bn,ji} - p_{bn,ij} - p_{bn,ji}\right) + \left(\hat{p}_{an,ij} - p_{an,ij}\right)\left(p_{bn,ij} + p_{bn,ji}\right)$ and Lemma B2, while the second term in eq. (B15) equals $o_p(1)$ by closely following the proof of Proposition 2 in Lin and Lee (2010).

Next, we consider

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{p}_{\Delta_n,ij}\hat{e}_{i,n}^2\hat{e}_{j,n}^2-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{p}_{\Delta_n,ij}e_{i,n}^2e_{j,n}^2\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij}-p_{\Delta_n,ij}\right)\left(\hat{e}_{i,n}^2\hat{e}_{j,n}^2-e_{i,n}^2e_{j,n}^2\right)+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}p_{\Delta_n,ij}\left(\hat{e}_{i,n}^2\hat{e}_{j,n}^2-e_{i,n}^2e_{j,n}^2\right)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij}-p_{\Delta_n,ij}\right)\left(\hat{e}_{i,n}^2\hat{e}_{j,n}^2-e_{i,n}^2e_{j,n}^2\right)+o_p(1)
\end{aligned} \tag{B16}
$$

where we closely follow the proof of Proposition 2 in Lin and Lee (2010) to obtain the last line, and the residuals can be decomposed as follows

$$
\begin{aligned}
\hat{e}_n &= \hat{S}_nY_n - X_n(W_n)\hat{\theta}_\beta - X_n(W_n,\hat{\lambda})\hat{\delta}_{\theta_\beta}\\
&= \left(\hat{S}_n - S_n\right)Y_n - X_n(W_n)\left(\hat{\theta}_\beta-\theta_{\beta_0}\right) - \left[X_n(W_n,\hat{\lambda}) - X_n(W_n,\lambda_0)\right]\hat{\delta}_{\theta_\beta}\\
&\quad -X_n(W_n,\lambda_0)\left(\hat{\delta}_{\theta_\beta}-\delta_{\theta_{\beta_0}}\right)+e_n\\
&= e_n + \check{b}_n + \check{c}_n
\end{aligned}
$$

where $\check{b}_n = \left(\hat{S}_n - S_n\right)S_n^{-1}e_n$ and

$$
\begin{aligned}
\check{c}_n = &\left(\hat{S}_n - S_n\right)S_n^{-1}\left[X_n(W_n)\theta_{\beta_0} + X_n(W_n,\lambda_0)\delta_{\theta_{\beta_0}}\right] - \left[X_n(W_n,\hat{\lambda}) - X_n(W_n,\lambda_0)\right]\hat{\delta}_{\theta_\beta}\\
&-X_n(W_n)\left(\hat{\theta}_\beta-\theta_{\beta_0}\right) - X_n(W_n,\lambda_0)\left(\hat{\delta}_{\theta_\beta}-\delta_{\theta_{\beta_0}}\right).
\end{aligned} \tag{B17}
$$

From the proof of Lemma B2, we actuall can extend the convergence in spectral normal results into the convergence for the row-/column sum norm results: $\left\|W_n^-(\hat{\lambda}) - W_n^-(\lambda_0)\right\|_1 = o_p(1)$, $\left\|W_n(\hat{\lambda}) - W_n(\lambda_0)\right\|_\infty = o_p(1)$, and $\left\|W_{n,\hat{\lambda}}^- - W_{n,\lambda_0}^-\right\|_1 =$

$o_p(1)$ and $\left\|W^{-}_{n,\hat{\lambda}} - W^{-}_{n,\lambda_0}\right\|_\infty = o_p(1)$. Hence, we obtain that $\check{b}_n = \bar{b}_n + o_p(1)$ and $\check{c}_n = \bar{c}_n + o_p(1)$ hold uniformly, where the result for $\check{c}_n$ requires all the elements of X_n to be uniformly bounded. [5]

Let $\hat{e}_{i,n}$, $e_{i,n}$, $\check{b}_{i,n}$, $\bar{b}_{i,n}$, $\check{c}_{i,n}$, and $\bar{c}_{i,n}$ denote the ith element of $\hat{e}_n$, e_n, $\check{b}_n$, $\bar{b}_n$, $\check{c}_n$, and $\bar{c}_n$, respectively. Then, $\hat{e}^2_{i,n} = \left(e_{i,n} + \check{b}_{i,n} + \check{c}_{i,n}\right)^2 = e^2_{i,n} + \check{b}^2_{i,n} + \check{c}^2_{i,n} + 2\left(e_{i,n}\check{b}_{i,n} + \check{c}_{i,n}e_{i,n} + \check{b}_{i,n}\check{c}_{i,n}\right)$. As $\hat{e}^2_{i,n}\hat{e}^2_{j,n} - e^2_{i,n}e^2_{j,n} = e^2_{j,n}\left(\hat{e}^2_{i,n} - e^2_{i,n}\right) + e^2_{i,n}\left(\hat{e}^2_{j,n} - e^2_{j,n}\right) + \left(\hat{e}^2_{j,n} - e^2_{j,n}\right)\left(\hat{e}^2_{i,n} - e^2_{i,n}\right)$, we can decompose the leading term in (B16) into three terms:

$$
\begin{aligned}
A_n &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right)\left(\hat{e}^2_{i,n}\hat{e}^2_{j,n} - e^2_{i,n}e^2_{j,n}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right)e^2_{j,n}\left(\hat{e}^2_{i,n} - e^2_{i,n}\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right)e^2_{i,n}\left(\hat{e}^2_{j,n} - e^2_{j,n}\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right)\left(\hat{e}^2_{j,n} - e^2_{j,n}\right)\left(\hat{e}^2_{i,n} - e^2_{i,n}\right) \\
&= A_{n1} + A_{n2} + A_{n3},
\end{aligned}
$$

where we define A_{n1}, A_{n2} and A_{n3} according to its order of appearance. For A_{n1}, we have

$$
\begin{aligned}
|A_{n1}| &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e^2_{j,n}\left|\hat{e}^2_{i,n} - e^2_{i,n}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e^2_{j,n}\left[3\left(\bar{b}^2_{i,n} + \bar{c}^2_{i,n}\right) + 2e^2_{i,n}\right] \\
&= \frac{3}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e^2_{j,n}\left(\bar{b}^2_{i,n} + \bar{c}^2_{i,n}\right) + o_p(1)
\end{aligned}
$$

by Lemma B2 and $\|P_{\varpi_n}\|_1 \leq M$. Letting $V_n = W_n S_n^{-1}\left[X_n(W_n)\theta_{\beta_0} + X_n(W_n, \lambda_0)\delta_{\theta_{\beta_0}}\right] = [v_{1,n}, \ldots, v_{n,n}]'$, we have $\max_{1\leq i\leq n}|v_{i,n}| = O(1)$ under Assumption (1.2) and the typical

[5] $\bar{b}_n = \hat{\delta}_{\alpha\alpha}\left(W_n^{-}(\hat{\lambda}) - W_n^{-}(\lambda_0)\right)S_n^{-1}e_n$ and $\bar{c}_n = \hat{\delta}_{\alpha\alpha}\left(W_n^{-}(\hat{\lambda}) - W_n^{-}(\lambda_0)\right)S_n^{-1}\left[X_n(W_n)\theta_{\beta_0} + X_n(W_n, \lambda_0)\delta_{\theta_{\beta_0}}\right]$.

element of $\bar{c}_n = \hat{\delta}_{\alpha\alpha} Z\left(\hat{\lambda}, \lambda_0\right) V_n$ is $\bar{c}_{i,n} = \hat{\delta}_{\alpha\alpha} \chi_i \left(\hat{\lambda}, \lambda_0\right) v_{i,n}$. It then follows that

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e_{j,n}^2 \bar{c}_{i,n}^2 \\
= & \frac{\hat{\delta}_{\alpha\alpha}^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e_{j,n}^2 \chi_i^2 \left(\hat{\lambda}, \lambda_0\right) v_{i,n}^2 \\
= & o_p(1).
\end{aligned}
$$

Moreover, letting $d_{ij,n}$ be the (i,j)th element of $W_n S_n^{-1}$ and denoting $V_n = W_n S_n^{-1} e_n = [v_{1,n}, \ldots, v_{n,n}]'$, we have $v_{i,n} = \sum_{l=1}^{n} d_{il,n} e_l$ and applying tedious but straightforward calculation gives

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e_{j,n}^2 \bar{b}_{i,n}^2 \\
= & \frac{\hat{\delta}_{\alpha\alpha}^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\hat{p}_{\Delta_n,ij} - p_{\Delta_n,ij}\right| e_{j,n}^2 \chi_i^2 \left(\hat{\lambda}, \lambda_0\right) \left(\sum_{l=1}^{n} d_{il,n} e_l\right)^2 \\
= & o_p(1).
\end{aligned}
$$

because $\|W_n S_n^{-1}\|_1 \le M$ under Assumption (1.2). Therefore, we obtain $A_{n1} = o_p(1)$. Similarly, we have $A_{n2} = o_p(1)$ and $A_{n3} = o_p(1)$.
Hence, we have $n^{-1}\left[\operatorname{tr}\left(\hat{\Gamma}_n \hat{P}_{an} \left(\hat{\Gamma}_n \hat{P}_{bn}\right)^s\right) - \operatorname{tr}\left(\Gamma_n P_{an} \left(\Gamma_n P_{bn}\right)^s\right)\right] = o_p(1)$ for $a, b = 1, 2, 3, 4$.

Next, we consider

$$
\begin{aligned}
& \hat{Q}_n' \hat{\Gamma}_n \hat{Q}_n - Q_n' \Gamma_n Q_n \\
= & \left(\hat{Q}_n - Q_n\right)' \hat{\Gamma}_n \left(\hat{Q}_n - Q_n\right) + \left(\hat{Q}_n - Q_n\right)' \hat{\Gamma}_n Q_n \\
& + Q_n' \hat{\Gamma}_n \left(\hat{Q}_n - Q_n\right) + Q_n' \left(\hat{\Gamma}_n - \Gamma_n\right) Q_n
\end{aligned}
$$

where Q_n equals $\hat{Q}_n$ with $\hat{\theta}$ replaced with θ_0 and
$\hat{Q}_n - Q_n = \left[\Delta_{n1}, \Delta_{n2}, \Delta_{n3}, \Delta_{n4}, 0_{n \times (2k_1)}, X_n(W_n, \hat{\lambda}) - X_n(W_n, \lambda_0)\right]$ with

$$
\begin{aligned}
\Delta_{n1} &= W_n \hat{S}_n^{-1} [X_n(W_n) \hat{\theta}_\beta, X_n(W_n, \hat{\lambda}) \hat{\delta}_{\theta_\beta}] - W_n S_n^{-1} [X_n(W_n) \theta_{\beta_0}, X_n(W_n, \lambda_0) \delta_{\theta_{\beta_0}}] \\
\Delta_{n2} &= W_{n,\hat{\lambda}}^{-} \hat{S}_n^{-1} [X_n(W_n) \hat{\theta}_\beta, X_n(W_n, \hat{\lambda}) \hat{\delta}_{\theta_\beta}] - W_{n,\lambda_0}^{-} S_n^{-1} [X_n(W_n) \theta_{\beta_0}, X_n(W_n, \lambda_0) \delta_{\theta_{\beta_0}}] \\
\Delta_{n3} &= W_n(\hat{\lambda}) \hat{S}_n^{-1} [X_n(W_n) \hat{\theta}_\beta, X_n(W_n, \hat{\lambda}) \hat{\delta}_{\theta_\beta}] - W_n(\lambda_0) S_n^{-1} [X_n(W_n) \theta_{\beta_0}, X_n(W_n, \lambda_0) \delta_{\theta_{\beta_0}}] \\
\Delta_{n4} &= W_n^{-}(\hat{\lambda}) \hat{S}_n^{-1} [X_n(W_n) \hat{\theta}_\beta, X_n(W_n, \hat{\lambda}) \hat{\delta}_{\theta_\beta}] - W_n^{-}(\lambda_0) S_n^{-1} [X_n(W_n) \theta_{\beta_0}, X_n(W_n, \lambda_0) \delta_{\theta_{\beta_0}}].
\end{aligned}
$$

Letting $\hat{q}_{i,n}$ and $q_{i,n}$ be the ith column of $\hat{Q}_n$ and Q_n, respectively, we have

$$\begin{aligned} n^{-1}\left(\hat{Q}_n - Q_n\right)' \hat{\Gamma}_n \left(\hat{Q}_n - Q_n\right) &= n^{-1}\sum_{i=1}^{n} \hat{e}_{i,n}^2 \left(\hat{q}_{i,n} - q_{i,n}\right)\left(\hat{q}_{i,n} - q_{i,n}\right)' \\ n^{-1}\left(\hat{Q}_n - Q_n\right)' \hat{\Gamma}_n Q_n &= n^{-1}\sum_{i=1}^{n} \hat{e}_{i,n}^2 \left(\hat{q}_{i,n} - q_{i,n}\right) q_{i,n}' \\ n^{-1} Q_n' \left(\hat{\Gamma}_n - \Gamma_n\right) Q_n &= n^{-1}\sum_{i=1}^{n} \left(\hat{e}_{i,n}^2 - e_{i,}^2\right) q_{i,n} q_{i,n}'. \end{aligned}$$

Using the results in Lemma B2 and the arguments made above, we can show that the three equations are all of order $o_p(1)$ element by element. That is, we have $n^{-1}\left(\hat{Q}_n'\hat{\Gamma}_n\hat{Q}_n - Q_n'\Gamma_n Q_n\right) = o_p(1)$.

To sum up, we have shown that $n^{-1}\left(\hat{\Omega}_n - \Omega_n\right) = o_p(1)$. This completes the proof of this proposition.

Proof of Theorem 2: From the generalized Schwartz inequality, we know that the optimal weighting matrix is $(n^{-1}\Omega_n)^{-1}$. Setting $a_n = (n^{-1}\Omega_n)^{-1/2}$, we have $a_0 = (\lim_{n\to\infty} n^{-1}\Omega_n)^{-1/2}$ exists under Assumption 6. Applying simple algebra yields

$$\begin{aligned} & \hat{g}_n(\theta)'\hat{\Omega}_n^{-1}\hat{g}_n(\theta) - g_n(\theta)'\hat{\Omega}_n^{-1}g_n(\theta) \\ = \; & g_n(\theta)'\hat{\Omega}_n^{-1}g_n(\theta) + 2\left[\hat{g}_n(\theta) - g_n(\theta)\right]'\hat{\Omega}_n^{-1}g_n(\theta) \\ & + \left[\hat{g}_n(\theta) - g_n(\theta)\right]'\hat{\Omega}_n^{-1}\left[\hat{g}_n(\theta) - g_n(\theta)\right]. \end{aligned}$$

Applying Lemma B2, we can show $\max_{\theta\in\Theta} n^{-1}\left\|\hat{g}_n(\theta) - g_n(\theta)\right\| = o_p(1)$, which implies $n^{-1}\max_{\theta\in\Theta}\left\|\hat{g}_n(\theta)'\hat{\Omega}_n^{-1}\hat{g}_n(\theta) - g_n(\theta)'\hat{\Omega}_n^{-1}g_n(\theta)\right\| = o_p(1)$.

In addition, we have $n^{-1}g_n(\theta)'\hat{\Omega}_n^{-1}g_n(\theta) = n^{-1}g_n(\theta)'\Omega_n^{-1}g_n(\theta) + n^{-1}g_n(\theta)'(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta) = n^{-1}g_n(\theta)'\Omega_n^{-1}g_n(\theta) + o_p(1)$ by Proposition 2.2, because $\max_{\theta\in\Theta}\left\|n^{-1}g_n(\theta) - n^{-1}E\left[g_n(\theta)\right]\right\| = o_p(1)$ and $\max_{\theta\in\Theta}\left\|n^{-1}E\left[g_n(\theta)\right]\right\| = O(1)$. Therefore, minimizing $\hat{g}_n(\theta)'\hat{\Omega}_n^{-1}\hat{g}_n(\theta$ w.r.t. $\theta\in\Theta$ is equivalent to minimizing $n^{-1}g_n(\theta)'\Omega_n^{-1}g_n(\theta)$ over $\theta\in\Theta$. Following the proof of Theorem 2.14, we therefore obtain

$$\begin{bmatrix} \sqrt{n}\left(\hat{\theta}^{**} - \theta_0^{**}\right) \\ n^{\frac{1}{2}-a}\left(\hat{\lambda} - \lambda_0\right) \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0}, \Sigma_n^*\right) \tag{B18}$$

where $\Sigma_n^* = (\lim_{n\to\infty}\frac{1}{n}H_n'\Lambda_n\Omega_n^{-1}\Lambda_n H_n)^{-1}$. This completes the proof of this theorem.

Appendix C

Lemma C1 *Under Assumption 1, $\vartheta_1 \neq \vartheta_2$ and $H_0 : a_1 = a_2 = 0$, the following hold*

1. $|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}| = O_p(E(z_i x_i'|q{=}\gamma_0)n^{-1+2\alpha})$.
2. $|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \gamma_0\}| = O_p(E(z_i x_i'|q{=}\gamma_0)n^{-1+2\alpha})$.
3. $|n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}| = O_p(\sqrt{E(z_i z_i'|q{=}\gamma_0)/n}n^{-1+2\alpha})$.
4. $|n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \gamma_0\}| = O_p(\sqrt{E(z_i z_i'|q{=}\gamma_0)/n}n^{-1+2\alpha})$.

where $z_i = x_i$ or w_i.

Proof of Lemma C1.

1. Note that
   $|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}| \leq n^{-1}\sum_{i=1}^n |z_i x_i'| 1\{\gamma_0 - |\hat{\gamma} - \gamma_0| < q_i < \gamma_0 + |\hat{\gamma} - \gamma_0|\}$.

   Given that $\hat{\gamma} = \gamma_0 + O_p\left(n^{-1+2\alpha}\right)$, for any small $\epsilon > 0$, there exists a constant M and an integer N_ϵ such that for any $n > N_\epsilon$, $\Pr\left\{|\hat{\gamma} - \gamma_0| > Mn^{-1+2\alpha}\right\} \leq \epsilon$. Hence, for any $n > N_\epsilon$ and any finite $\tilde{M} > 0$ such that

$$
\begin{aligned}
&\Pr\left\{|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}| > \tilde{M}E(z_i x_i'|q_i = \gamma_0)n^{-1+2\alpha}\right\} \\
&= \\
&\Pr\left\{\left\{|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}| > \tilde{M}E(z_i x_i'|q_i = \gamma_0)n^{-1+2\alpha}\right\}\right. \\
&\qquad\qquad \left.\cap\left\{|\hat{\gamma} - \gamma_0| \leq Mn^{-1+2\alpha}\right\}\right\} \\
&+ \\
&\Pr\left\{\left\{|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}| > \tilde{M}E(z_i x_i'|q_i = \gamma_0)n^{-1+2\alpha}\right\}\right. \\
&\qquad\qquad \left.\cap\left\{|\hat{\gamma} - \gamma_0| > Mn^{-1+2\alpha}\right\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \Pr\left\{ n^{-1}\sum_{t=1}^{n} |z_i x_i'| 1\{\gamma_0 - Mn^{-1+2\alpha} < q_i < \gamma_0 + Mn^{-1+2\alpha}\} > \tilde{M}E(z_i x_i'|q{=}\gamma_0) n^{-1+2\alpha} \right\} \\
&\leq \Pr\left\{ n^{-1}\sum_{t=1}^{n} |z_i x_i'| 1\{\gamma_0 - Mn^{-1+2\alpha} < q_i < \gamma_0 + Mn^{-1+2\alpha}\} > \tilde{M}E(z_i x_i'|q{=}\gamma_0) n^{-1+2\alpha} \right\} \\
&\quad + \Pr\left\{ |\hat{\gamma} - \gamma_0| > Mn^{-1+2\alpha} \right\} \\
&\leq \frac{n^{-1}\sum_{t=1}^{n} E\left[|z_i x_i'| 1\{\gamma_0 - Mn^{-1+2\alpha} < q_i < \gamma_0 + Mn^{-1+2\alpha}\}\right]}{\tilde{M}E(z_i x_i'|q{=}\gamma_0) n^{-1+2\alpha}} + \epsilon \\
&= \frac{2E\left[|z_1 x_1'| \, |\mathbf{q}_1 = \breve{\gamma}\right] f_q(\breve{\gamma}) M}{\tilde{M}E(z_i x_i'|q{=}\gamma_0)} + \epsilon .
\end{aligned}
$$

The second inequality is obtained after applying Markov's inequality and last equality from mean value theorem to obtain the last equation with $\breve{\gamma}$ lies between $\gamma_0 - Mn^{-1+2\alpha}$ and $\gamma_0 + Mn^{-1+2\alpha}$. Hence $|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\} 1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\}| = O_p(E(z_i x_i'|q{=}\gamma_0) n^{-1+2\alpha})$.

2. Similarly, $|n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \hat{\gamma}\} 1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n z_i x_i' 1\{q_i \leq \gamma_0\}| = O_p(E(z_i x_i'|q{=}\gamma_0) n^{-1+2\alpha})$.

3. Let the partition $[\gamma_0 - Mn^{-1+2\alpha},\ \gamma_0 + Mn^{-1+2\alpha}] = \cup_{k=1}^{N_\epsilon - 1}[\gamma_k, \gamma_{k+1}) \cup [\gamma_{N\epsilon}, \gamma_{N_\epsilon+1}]$ into N_ϵ non-overlapping intervals with equal length $\epsilon = 2Mn^{-1+2\alpha}/N_\epsilon$.

   Then, we have

$$
\begin{aligned}
&\max_{\hat{\gamma} \in [\gamma_0 - Mn^{-1+2\alpha}, \gamma_0 + Mn^{-1+2\alpha}]} |n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\} 1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}| \\
\leq\ & \max_{|\gamma' - \gamma| < \epsilon} |n^{-1}\sum_i^n e_i z_i \left[1\{q_i \leq \gamma'\} - 1\{q_i \leq \gamma\}\right] 1\{q_i > \gamma_0\}| \\
&+ \sum_{k=1}^{N_\epsilon+1} |n^{-1}\sum_{i=1}^n e_i z_i 1\{q_i \leq \gamma_k\} 1\{q_i > \gamma_0\}| \\
=\ & O_p\left(\sqrt{E\left(z_i' z_i | q_i = \gamma_0\right)/n\epsilon}\right) + O_p\left(N_\epsilon n^{-1+2\alpha}\sqrt{E\left(z_i' z_i | q_i = \gamma_0\right)/n}\right) \\
=\ & O_p\left(\sqrt{E\left(z_i' z_i | q_i = \gamma_0\right)/n}\left(\epsilon + N_\epsilon n^{-1+2\alpha}\right)\right) = O_p\left(n^{-1+2\alpha)}\sqrt{E\left(z_i' z_i | q_i = \gamma_0\right)/n}\right)
\end{aligned}
$$

   for any finite N_ϵ. It follows,

$$\Pr\left\{|n^{-}\sum_{i}^{n} e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_{i}^{n} e_i z_i' 1\{q_i \leq \hat{\gamma}\}| > \tilde{M} b_n\right\}$$

$$=$$

$$\Pr\left\{\left\{|n^{-}\sum_{i}^{n} e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_{i}^{n} e_i z_i' 1\{q_i \leq \hat{\gamma}\}| > \tilde{M} b_n\right\} \cap \left\{|\hat{\gamma} - \gamma_0| \leq M n^{-1+2\alpha}\right\}\right\}$$

$$+$$

$$\Pr\left\{\left\{|n^{-}\sum_{i}^{n} e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_{i}^{n} e_i z_i' 1\{q_i \leq \hat{\gamma}\}| > \tilde{M} b_n\right\} \cap \left\{|\hat{\gamma} - \gamma_0| > M n^{-1+2\alpha}\right\}\right\}$$

$$\leq \Pr\left\{\max_{|\hat{\gamma}-\gamma_0| \leq \boldsymbol{M}_{\epsilon} n^{-1+2\alpha}} |n^{-1}\sum_{t=1}^{n} e_i z_i' 1\{q_t \leq \hat{\gamma}\}1\{q_t > \gamma_0\}| > \tilde{M} b_n\right\}$$
$$+ \Pr\left\{|\hat{\gamma} - \gamma_0| > M n^{-1+2\alpha}\right\} = \epsilon + \epsilon = 2\epsilon,$$

where $b_n = \sqrt{E(z_i z_i'|q_{=}\gamma_0)/n} n^{-1+2\alpha}$. Hence, $|n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}| = O_p(\sqrt{E(z_i z_i'|q_{=}\gamma_0)/n} n^{-1+2\alpha})$.

4. Similarly,
$|n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \hat{\gamma}\}1\{q_i \leq \gamma_0\} - n^{-}\sum_i^n e_i z_i' 1\{q_i \leq \gamma_0\}| = O_p(\sqrt{E(z_i z_i'|q_{=}\gamma_0)/n} n^{-1+2\alpha})$.

Proof of Proposition 3.1. Let the model

$$y_{i,n} = \begin{cases} a_1 w_i + \vartheta_1' x_i + e_i, & q_i \leq \gamma \\ a_2 w_i + \vartheta_2' x_i + e_i, & q_i > \gamma \end{cases}$$

which is equivalent to

$$y_i = \vartheta' x_i + a w_i + \delta_1 x_i 1\{q_i \leq \gamma\} + \delta_2 w_i 1\{q_i \leq \gamma\} + e_i$$

where $\vartheta = \vartheta_2$, $a = a_2$, $\delta_1 = \vartheta_1 - \vartheta_2$ and $\delta_2 = a_1 - a_2$.

Under the null hypothesis $H_o : a_1 = a_2 = 0$ or $(H_o : \delta_2 = a = 0)$ the model reduces to

$$y_i = \vartheta' x_i + \delta_1 x_i 1\{q_i \leq \gamma\} + e_i \tag{C1}$$

When $\vartheta_1 \neq \vartheta_2$ ($\delta_1 = 0$), we have the standard threshold regression model and from Hansen (2000), $n^{1-2\alpha}(\hat{\gamma} - \gamma_o) \xrightarrow{d} \omega T$, where $\omega = \frac{C'VC}{(C'DC)^2 f}$, $T = arg\max_{-\infty<r<\infty}[-\frac{1}{2}r + W(r)]$.

When $\vartheta_1 = \vartheta_2$, model (C1) becomes $y_i = \vartheta' x_i + e_i$. Define the matrix Q_γ, of stacked elements $[x_i 1\{q_i \leq \gamma\}, x_i 1\{q_i > \gamma\}]$, $M_\gamma = I_n - Q_\gamma(Q'_\gamma Q_\gamma)^{-1} Q_\gamma$ and $\lambda = F_q(\gamma)$. Then

$$\begin{aligned} y' M_\gamma y &= y'(I_n - Q_\gamma(Q'_\gamma Q_\gamma)^{-1} Q_\gamma) y \\ &= e'(I_n - Q_\gamma(Q'_\gamma Q_\gamma)^{-1} Q_\gamma) e \\ &= e'e - e' Q_\gamma (Q'_\gamma Q_\gamma)^{-1} Q_\gamma e \end{aligned}$$

From Law of Large Numbers, $n^{-1} e'e \xrightarrow{a.s} \sigma^2$ and

$$\begin{aligned} & e' Q_\gamma (Q'_\gamma Q_\gamma)^{-1} Q_\gamma e \Rightarrow B_u(1,\lambda)' B_u(1,\lambda) \\ + \ & [E_{xx}^{1/2} B_u(1) - E_{xx,\gamma}^{1/2} B_u(1,\lambda)]'[E(x_i' x_i) - E(x_i' x_i 1\{q_i \leq \gamma\})]^{-1}[E_{xx}^{1/2} B_u(1) - E_{xx,\gamma}^{1/2} B_u(1,\lambda)] \end{aligned}$$

where $E_{xx} = E(x_i' x_i)$ and $E_{xx,\gamma} = E(x_i' x_i 1\{q_i \leq \gamma\})$, because

$$\begin{aligned} e' Q_\gamma (Q'_\gamma Q_\gamma)^{-1} Q_\gamma e &= \sum_i^n x_i' e_i 1\{q_i \leq \gamma\} (\sum_i^n x_i x_i' \{q_i \leq \gamma\})^{-1} \sum_i^n x_i e_i 1\{q_i \leq \gamma\} \\ &+ \sum_i^n x_i' e_i 1\{q_i > \gamma\} (\sum_i^n x_i x_i' \{q_i > \gamma\})^{-1} \sum_i^n x_i e_i 1\{q_i > \gamma\} \end{aligned}$$

and from Theorem 1 of Hansen and Caner (2001), $[n^{-1/2} E_x x^{-1/2} \sum_i^{ns} x_i e_i, n^{-1/2} E_{xx,\gamma}^{-1/2} \sum_i^{ns} x_i e_i 1\{q_i > \gamma\}] \Rightarrow [B_u(s), B_u(s,\lambda)]$. Now, denote $\lambda^* = F_q(\gamma^*)$ and $\hat{\lambda} = F_q(\hat{\gamma})$. Then

$$\begin{aligned} \hat{\lambda} \Rightarrow \lambda^* &= argmax B_u(1,\lambda)' B_u(1,\lambda) \\ &+ [E_{xx}^{1/2} B_u(1) - E_{xx,\gamma}^{1/2} B_u(1,\lambda)]'[E(x_i' x_i) - E(x_i' x_i 1\{q_i \leq \gamma\})]^{-1}[E_{xx}^{1/2} B_u(1) - E_{xx,\gamma}^{1/2} B_u(1,\lambda)] \end{aligned}$$

Proof of Proposition 3.2. Let the Wald statistic

$$W_n(\hat{\gamma}) = \frac{(R'\hat{\beta})'(R'(X(\hat{\gamma})' X(\hat{\gamma}))^{-1} R)^{-1}(R'\hat{\beta})}{s^2} \tag{C2}$$

where R is a selection matrix, $\hat{\gamma}$ is the γ estimated under the null, s^2 is the residual variance calculated under the alternative, $\hat{\beta} = [\vartheta, a, \delta_1, \delta_2]$ and $X(\gamma)$ is the matrix of stacked elements $[x_i 1\{q_i \leq \gamma\}, w_i 1\{q_i \leq \gamma\}, x_i 1\{q_i > \gamma\}, w_i 1\{q_i > \gamma\}]$.

Let $X(\gamma) = [X_-(\gamma), X_+(\gamma)]$, where $X_-(\gamma)$ is the matrix of stacked elements $[x_i 1\{q_i \leq \gamma\}, w_i 1\{q_i \leq \gamma\}]$and $X_+(\gamma)$ is the matrix of stacked elements $[x_i 1\{q_i > \gamma\}, w_i 1\{q_i > \gamma\}]$.

Hence,

$$[X(\gamma)'X(\gamma)]^{-1} = \left(\begin{bmatrix} X_-(\gamma)' \\ X_+(\gamma)' \end{bmatrix} \begin{bmatrix} X_-(\gamma)' & X_+(\gamma)' \end{bmatrix} \right)^{-1} = \begin{bmatrix} X_-(\gamma)'X_-(\gamma) & 0 \\ 0 & X_+(\gamma)'X_+(\gamma) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} (X_-(\gamma)'X_-(\gamma))^{-1} & 0 \\ 0 & (X_+(\gamma)'X_+(\gamma))^{-1} \end{bmatrix}$$

From Lemma 1, Hansen (1996)

$$n^{-1}(X(\gamma)'X(\gamma) \xrightarrow{p} \begin{bmatrix} (E[x_i'^{*}x_i^{*}1\{q_i \leq \gamma\}])^{-1} & 0 \\ 0 & E[x_i'^{*}x_i^{*}1\{q_i > \gamma\}])^{-1} \end{bmatrix},$$

Hence $R'(X(\hat{\gamma})'X(\hat{\gamma}))^{-1}R)^{-1} = O_p(n)$.

Under the null $y = \tilde{X}_-(\gamma)\vartheta_1 + \tilde{X}_+(\gamma)\vartheta_2 + e$, where $\tilde{X}_-(\gamma)$ is the matrix whose $i'th$ row is $x_i 1\{q_i \leq \gamma\}$ and $\tilde{X}_+(\gamma)$ the matrix whose $i'th$ row is $x_i 1\{q_i > \gamma\}$. Note that $[\hat{a_1}, \hat{a_2}]' = (R'\hat{\beta})'$ and define W^* the matrix of stacked elements $[w_i 1\{q_i \leq \gamma\}, w_i 1\{q_i > \gamma\}]$. Then

$$[\hat{a_1}, \hat{a_2}] = (W^{*'}M_\gamma W^*)^{-1}W^{*'}M_\gamma y = (W^{*'}M_\gamma W^*)^{-1}W^{*'}M_\gamma(X_-\tilde{(}\gamma)\vartheta_1 + X_+\tilde{(}\gamma)\vartheta_2 + e)$$

First, we consider the case where $\vartheta_1 = \vartheta_2$. Note that, $R'(X(\hat{\gamma})'X(\hat{\gamma}))^{-1}R)^{-1} \xrightarrow{d} R'(X(\gamma^*)'X(\gamma^*))^{-1}R)^{-1} = O_p(n)$ from continuous mapping theorem and $\hat{\gamma} \xrightarrow{d} \gamma^*$. Then $y = \tilde{X}\vartheta + e$, where $\tilde{X}$ is the matrix whose $i'th$ row is x_i. Then

$$\hat{a}, = (W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}y = (W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}(\tilde{X}\vartheta + e)$$

Since $\tilde{X}$ lies in the space spanned by $Q_{\hat{\gamma}}(Q_{\hat{\gamma}}'Q_{\hat{\gamma}})^{-1}Q_{\hat{\gamma}}$ hence $(W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}\tilde{X}\vartheta = 0$. Furthermore, $n^{-1}(W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}e \xrightarrow{d} n^{-1}(W^{*'}M_{\gamma^*}W^*)^{-1}W^{*'}M_{\gamma^*}e$ from continuous mapping theorem and $\hat{\gamma} \xrightarrow{d} \gamma^*$, which in turn converges to a zero mean normal distribution when e_i is independent of x_i. Hence, $W_n(\hat{\gamma}) \xrightarrow{d} \chi_2^2$.

Next, we consider the case where $\vartheta_1 \neq \vartheta_2$. Similarly $R'(X(\hat{\gamma})'X(\hat{\gamma}))^{-1}R)^{-1} = O_p(n)$. When $\vartheta_1 \neq \vartheta_2$,

$$[\hat{a}_1, \hat{a}_2] = (W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}y = (W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}(\tilde{X}_-(\gamma_0)\vartheta_1 + \tilde{X}_+(\gamma_0)\vartheta_2 + e)$$

Note that from Proposition 1, $n^{1-2\alpha}(\hat{\gamma}-\gamma_0) \xrightarrow{d} \omega T$. Hence, $\hat{\gamma} = \gamma_0 + O_p(n^{-1+2\alpha})$. Therefore, from Lemma 1 we obtain that $(W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}(\tilde{X}_-(\gamma_0)\vartheta_1 + \tilde{X}_+(\gamma_0)\vartheta_2 = O_p(n^{-1+2\alpha})$ and $(W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}y = (W^{*'}M_{\hat{\gamma}}W^*)^{-1}W^{*'}M_{\hat{\gamma}}e$ converges to a zero mean normal distribution. Hence, $W_n(\hat{\gamma}) \xrightarrow{d} \chi_2^2$.