REGRESSION-ADJUSTED ESTIMATION OF QUANTILE TREATMENT EFFECTS UNDER COVARIATE-ADAPTIVE RANDOMIZATIONS

By

Liang Jiang, Peter C.B. Phillips, Yubo Tao and Yichong Zhang

COWLES FOUNDATION PAPER NO. 1858

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2023

http://cowles.yale.edu/

Challenge

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/mcha20

John Maynard Keynes Narrates the Great Depression: His Reports to the Philips Electronics Firm

Robert W. Dimand & Bradley W. Bateman

To cite this article: Robert W. Dimand & Bradley W. Bateman (26 Oct 2023): John Maynard Keynes Narrates the Great Depression: His Reports to the Philips Electronics Firm, Challenge, DOI: 10.1080/05775132.2023.2272544

To link to this article: https://doi.org/10.1080/05775132.2023.2272544

CHALLENGE
https://doi.org/10.1080/05775132.2023.2272544

John Maynard Keynes Narrates the Great Depression: His Reports to the Philips Electronics Firm†

Robert W. Dimand and Bradley W. Bateman

ABSTRACT
In October 1929, the Dutch electronics firm Philips approached John Maynatd Keynes to write confidential reports on the state of the British and world economies, which he did from January 1930 to November 1934, at first monthly and then quarterly. These substantial reports (Keynes's November 1931 report was twelve typed pages) show Keynes narrating the Great Depression in real time, as the world went through the US slowdown after the Wall Street crash, the Credit-Anstalt collapse in Austria, the German banking crisis (summer 1931), Britain's departure from the gold exchange standard in August and September 1931, the US banking crisis leading to the Bank Holiday of March 1933, the London Economic Conference of 1933, and the coming of the New Deal. This series of reports has not been discussed in the literature, though the reports and surrounding correspondence are in the Chadwyck-Healey microfilm edition of the Keynes Papers. We examine Keynes's account of the unfolding events of the early 1930s, his insistence that the crisis would be more severe and long-lasting than most observers predicted, and his changing position on whether monetary policy would be sufficient to promote recovery and relate his reading of contemporary events to his theoretical development.

Introduction

On October 23, 1929, just as Wall Street began to crash[1] and the world economy moved into exceptionally interesting times, Dr. H. F. van Walsem, counsel and secretary to the Dutch electronics firm N. V. Philips Gloeilampenfabrieken[2], wrote to "J. M. Keynes, Esq., C.B. Cambridge" asking him to write a monthly letter to the firm's Economic Intelligence Service about the state of the British economy and the world economy. John Maynard Keynes's letters to Philips, monthly from January 1930 to November 1931 and then, because of budget cuts to Philips's Economic Intelligence Service, quarterly from February 1932 to November 1934, show Keynes narrating the events of the Great Depression as they occurred, and reveal his perception of the convulsions of the

CONTACT Robert W. Dimand rdimand@brocku.ca, robert.dimand@yale.edu Department of Economics, Brock University, St. Catharines, Ontario L2S 3A1, Canada.

(Fall 2023: Visiting Professor of Economics, Yale University)
Robert W. Dimand is a professor of economics at Brock University, Canada, and a visiting professor at Yale University, and author of *The Origins of the Keynesian* and an editor of *The Elgar Companion to John Maynard Keynes*.
Bradley W. Bateman is president emeritus of Randolph College, Virginia, coauthor of *Capitalist Revolutionary: John Maynard Keynes* and an editor of *The Cambridge Companion to Keynes*.

world economy as he wrote his *General Theory of Employment, Interest and Money* (1936). This substantial body of Keynes's commentary on economic fluctuations (the November 1931 letter alone is twelve typed, double-spaced pages) has hitherto been neglected in the literature on Keynes. Keynes's reports and the associated correspondence, preserved in the Keynes Papers at King's College, Cambridge, are included in the 1993 Chadwyck-Healey microfilm edition of the Keynes Papers (section BM/5 Memoranda Exchanged with Business Houses), but the expense of this edition (which was sold only as a complete set of 170 reels of microfilm, priced at £9,700 or $17,000, plus $175 for a hardcover catalogue, Cox 1993) meant that only a few copies were sold. According to the WorldCat catalogue, there are five sets in libraries in the United States (Library of Congress, Harvard, Yale, Ohio State, and University of Texas at El Paso), two in Great Britain (Universities of Oxford and Sheffield), one in Canada (Victoria University in the University of Toronto) and a few in Germany (Göttingen), Italy and elsewhere but surprisingly little use has been made even of these copies of Keynes's letters to N. V. Philips. Neither Moggridge (1992) nor Skidelsky (1983–2000, 2003), major biographies of Keynes by the authors who know the Keynes Papers best, mentions Keynes's reports to Philips (but Backhouse and Bateman 2011, 129, have a paragraph about Keynes's July 1930 report). As Jacqueline Cox (1995, 171) notes, the thirty volumes of Keynes's *Collected Writings* (1971–1989) include "only a third of the bulk classified as economic" in the Keynes Papers at King's and do not include Keynes's philosophical papers there, while "the personal papers were barely touched." Donald Moggridge (2006, 136–137) observes that "There has, inevitably, been heavier use of the Keynes Papers in King's College Cambridge, which have the advantage of being available elsewhere on microfilm, than, say, his papers in the National Archives or his correspondence with his publishers, the last of which reveals the risks of depending on the Cambridge collection alone." A vast amount of research has been done about Keynes and his economics, yet not all the relevant material has been explored (see Backhouse and Bateman 2006, Dimand and Hagemann 2019).

These reports reveal Keynes's reading of what was happening in the British and world economies through the first four years of the Great Depression, and provide the empirical counterpart to the record of Keynes's theoretical development in this period given by notes taken by students at Keynes's lectures from 1932 to 1935 (Rymes 1987, 1989, Dimand 1988, Dimand and Hagemann 2019). After the success of *The Economic Consequences of the Peace* (1919), Keynes no longer needed to be paid for lecturing, and so gave a single series of eight lectures each year, on the subject of whatever book he was writing at the time, so his lectures from 1932 to 1935 are in effect annual drafts of the book that became *The General Theory*. These lectures at Cambridge and the reports to N. V. Philips on what was happening in the economy provide theoretical and empirical supplements to Keynes's *Collected Writings* (1971–1989), respectively, in following Keynes's intellectual development in the Great Depression, from *A Treatise on Money* (1930) to *The General Theory* (1936). In Keynes's workload, his reports to Philips from 1930 to 1934 took the place of the London and Cambridge Economics Service Special Memoranda on commodity markets that he wrote from 1923 to 1930 (Keynes [1923–30] 1983, 267–647), which provided an empirical counterpart to his normal backwardation theory of futures contracts ([1923] 1983, 1930, Chapter 29).

Replying on October 31 to von Walsem's letter inviting him to write the monthly letter to the firm's Economic Intelligence Service, Keynes was "quite ready to discuss this proposal with one of your representatives" but wished to clarify "that there will be no question of the publication of the letters and that they will be purely for the information of your own people" – and that "it would not be practicable to me to undertake such work except in return for a somewhat substantial fee which might be higher than you would be willing to offer." On November 4, von Walsem assured him that the letters would not be published and "There are only two persons who, though not in our service, are closely related to our firm, who also receive a copy of our Intelligence Service which they, however, are bound to consider as absolutely confidential." He suggested £100 a year. On November 13, Keynes, having "considered your kind proposal in relation to the fees which I have received on previous occasions for somewhat analogous work," offered to undertake the task for an initial six months, for £150 a year[3]. Although Van Walsem had initially asked for the suggestion of other authors if Keynes preferred not take on the task at the suggested £100 a year, and Keynes equally pointedly offered to suggest such alternative authors if Philips did not care to pay £150 a year, Van Walsem accepted Keynes's terms for Philips on November 22: "We think it desirable that one of our gentlemen will see you in order to discuss some details in the first half of December next."

In the event two representatives of Philips (Messrs. Sannes and du Pré) met with Keynes for a discussion summarized "for good order's sake" by van Walsem on December 21, 1929 (by which time van Walsem had already received a December 18 note by Keynes on the Australian exchange position). He recorded agreement that Keynes's monthly letter would treat "some important factor in the development of the British economic situation and give your opinion as to its effects on trade in general and on our business in particular. Also you will draw our attention to important events in the domains especially interesting us, in so far as these come to your knowledge … Whenever you think it necessary you will give us your views on the situation in different parts of the British Empire or eventually of other countries. If possible we shall suggest [to] you special points to be considered in your letters." Von Walsem wrote again on June 21, 1930 to confirm "that the arrangement has given us full satisfaction so that we are willing to continue on the same terms" and enclosed a cheque for 75 pounds. The arrangement also satisfied Keynes; he wrote on January 1, 1931, that "I have enjoyed preparing the letters." Keynes's letters balanced opinions about trade in general with observations about matters affecting Philips more specifically. Thus on January 11, 1930, Keynes stated that "The Factory capacity for Radio Sets seems to have become quite appalling during 1929" before proceeding more generally "to take this opportunity of emphasizing the anxiety which is felt here about the Australian position … I think that Australia may have more difficulties with her balance of trade during the coming year than the Argentine."[4]

The Slump of 1930: Investment, Debts and Deflation

Keynes's April 1930 letter suggested that, although a general improvement had not yet arrived, "there are a fair number of indications that we may be somewhere in the neighborhood of the bottom point." In particular, "the continuance of cheap money, and even more the expectation of such continuance, is bound to be effective in the situation in the course of a few months," but the effect on employment would be slower than on business feeling and the Stock Exchange and "it would not be surprising to see British unemployment figures go on mounting even to the neighborhood of 2,000,000 up to the end of this calendar year. … The effect of many rationalization schemes now in train will be for some time to come to improve profits rather than employment." With a large amount of Australian gold en route to the Bank of England, "there is less anxiety about the British exchange position than there has been for a very considerable time past" and Keynes expected the creation of the Bank for International Settlements to have a positive effect on confidence, a foreshadowing of his emphasis at Bretton Woods on the importance of designing appropriate international monetary institutions. Keynes doubted that the Federal Reserve Board would reverse its cheap money policy "until business and employment in the United States is a great deal better than it is now." This emphasis on expectations would be characteristic of Keynes's *General Theory* (although equally in line with Irving Fisher's quantity theoretic concern with expected inflation), as is the measurement of the ease of monetary policy by the cheapness of money, that is, by low nominal interest rates. Because nominal interest rates (especially short-term rates such as the Treasury Bill rate) were very low in a period of deflation, the Federal Reserve Board continued to view monetary conditions as easy throughout what Milton Friedman and Anna Schwartz (1963) later termed the "Great Contraction" of the US money supply (during which the monetary base increased, but not by enough to offset the rise in currency/deposit and reserve/deposit ratios), despite Fisher drawing the attention of his former student, Federal Reserve Governor Eugene Meyer, to the statistics on the shrinkage of the money supply, the sum of currency and demand deposits (Cargill 1992, Dimand 2019).

On June 24, 1930, H. du Pré emphasized that, "In reply to your remarks about the character of your monthly letters, we assure you that we leave it entirely to you to judge in each case which are the topics which are most worth being discussed by you." Nonetheless, "There is one question upon which we particularly should like to have your opinion." Keynes's monthly letters had repeatedly stated that recovery depended on the bond market becoming more active, with new loans being used not just for the refunding of floating debt but for new productive investment. "But on the other hand these last months many articles in the economic press" saw excessive capacity in many industries; "in other words that the world has first to grow into a productive apparatus which is too big for immediate needs. If this should be true, can a renewed investment-activity soon be hoped for, and if it soon comes, would it really do good? Of course there would be less unemployment in a number of industries; but would not prices of consumptive commodities, and so cost of living, rise? And especially it might turn out after some time, that the new activity has only added to the – supposed – actual over-investment, so that the disequilibrium would only be greater. It may of course be that entirely new industries are going to take the lead, but we do not yet see any that are very likely to do so. We should be much obliged if you would solve this puzzle for us or at least give your views on the pretended overcapacity and its probable effects on future developments in your next letter." This letter sheds light on the audience for Keynes's reports in the secretariat of N. V. Philips: not just salesmen looking for tips about the market for radio sets in Great Britain or elsewhere, but thoughtful businessmen pondering sophisticated economic issues such as the dual nature of productive investment in creating demand while increasing capacity (a problem to which the warranted growth rate of Harrod 1939 was an attempted solution).

In his July 1930 letter (seven typed pages, plus a six-page note on the bond market), Keynes warned that "it is now fully clear the world is in the middle of an international cyclical depression of unusual severity … a depression and a crisis of major dimensions … I believe that the prevailing opinion in the United States is still not pessimistic enough and is relying too much on a recovery in the early autumn, an event which is, in my opinion, most improbable. Nothing is more difficult than to predict the date of recovery. But all previous experience would show that a depression on this scale is not something from which the recovery comes suddenly or quickly." He felt that "The optimism of Wall Street and the hoarding tendencies of France may prevent any real recovery of the International Loan Market this year" and considered whether this might lead to "a psychological atmosphere in which really drastic scientific measures will be taken by Great Britain and the United States in conjunction to do what is humanly possible to cause a turn of the tide next spring. But one is traveling here into the realm of the altogether uncertain and unpredictable." In contrast, the Harvard Economic Society (founded by Harvard economics professors Charles J. Bullock and Warren Persons) stated in its weekly letter on June 28, 1930, that "irregular and conflicting movements of business should soon give way to sustained recovery" and on July 19 that "untoward elements have operated to delay recovery but the evidence nevertheless points to substantial improvement" (quoted by Galbraith 1961, 150, see also Walter Friedman 2014).

Responding to du Pré's query, Keynes reiterated that recovery would be preceded by "a substantial fall in the long-period rate of interest … leading in due course to the recovery of investment." But now he explained that he was not thinking of investment in manufacturing industry, "the world's capacity for which is probably quite ample for the present." Even at the highest estimate, the total cost of bringing Britain's industrial plant up to date "would not use up the country's savings for more than, say, three months. Moreover, when expected profits are satisfactory the rate of expenditure by manufacturing industry in fixed plant is not very sensitive to the rate of interest."

"On the other hand," in contrast to manufacturing, "the borrowing requirements for building, transport and public utilities are not only on a far greater scale, but are decidedly sensitive to the rate of interest. If I were to put my finger on the prime trouble to-day, I should call attention to the very high rate of interest for long-term borrowers … the long-term rate of interest is higher to-day than it has been in time of peace for a very long time past. When, at the same time, there is a big business depression and prices are falling, it is not surprising that new enterprise is kept back at the present level of interest." He drew attention to "those who might be called distress borrowers, that is say countries which have an urgent need for borrowing to pay off existing debts, and are consequently ready to pay a very high rate of interest," citing prospective Austrian, Hungarian and Australian loans on the London bond market, and remarked that "the effect of the German Loan has been to supply the French Treasury with funds, which it has withdrawn from the French market and is keeping unemployed in the

Bank of France." Keynes's July 1930 letter (discussed briefly by Backhouse and Bateman 2011, 129) illuminates both his analysis of the present situation and the role of investment in his economics. His distinction between investment in manufacturing, responsive to expected profit rather than interest rates, and interest-sensitive investment in construction, transport and public utilities clarifies his theory of investment. Increased investment was crucial for recovery of the world economy, and low long-term interest rates were necessary for high levels of investment in construction, transport and public utilities, the largest part of investment (even if manufacturing investment depended more on expected profits). In regard to the current situation, Keynes explained the forces getting long-term interest rates high even when prices were falling and short-term interest rates were low, but felt that "progress has been made toward getting the necessitous borrowers out of the way." On the immediate practical level, Keynes's distinction between the determinants of the two categories of investment dealt with du Pré's question of how low long-term interest rates could stimulate investment given excess productive capacity in manufacturing. And yet, unlike Harrod (1939), Keynes's July 1930 letter did not come to grips with the theoretical point raised by du Pré, the dual character of investment in creating both demand and productive capacity.

Keynes's August 1930 letter dissented from the view widely held in the United States "even in responsible quarters, that we may expect an autumn recovery with some confidence … a good deal of the American optimism is based on analogies drawn from the date of recovery after the 1920-21 slump" (compare the Harvard Economic Society's statement on August 30 that "the present depression has about spent its force," quoted by Galbraith 1961, 150). He argued that "Too much emphasis cannot be laid on the really catastrophic character of the price falls of some of the principal raw materials since a year ago" (even larger than appeared from published index numbers, because those included a number of commodities subject to price controls), which "must profoundly affect the purchasing power of all overseas markets." Long-term interest rates remained high, reducing new capital investment. In contrast, Keynes considered general opinion about the British position to be "perhaps a little too pessimistic." Britain was already in a difficult position before the slump of 1929 and 1930, because of the 1925 return to the gold exchange standard at the prewar parity (over the eloquent protests of Keynes 1925). But the heavy unemployment in the slump was limited to textiles and heavy industry (iron and steel, coal, and shipbuilding), export-based sectors already hit by the return to gold at an overvalued exchange rate (in his December 1930 letter, Keynes stated that if textiles, iron and steel, and coal were omitted, there was practically no decline in the Index of Production from a year before and an improvement from two years before). Keynes explained that British unemployment statistics, when used in international comparisons, "probably overstate the case" since the British statistics included "a great many workers in definite employment, but working short time … It is even the case that workers taking their normal summer holidays are now included in the figures of the unemployed." According to *The Economist*, the aggregate profits of all British joint stock companies reporting their earnings in the first half of 1930 "were not only greater than in the previous year, but were larger than in any previous year. This was partly due to the prosperity of British Oil Companies operating abroad, but by no means wholly." Nor did Keynes share the worries of financial opinion in London (and so some extent his own previous letter to Philips) about "the constant dribble of gold to France."

In Keynes's September 1930 letter to Philips, he was "still of the opinion that real recovery is a long way off. But at the same time it seems to me not unlikely that we are at, or near, the lowest point … It is time, therefore, to cease to be a 'bear', even if it is not yet time to be a 'bull'." His February 1931 letter began, "Glancing through the letters of previous months, I find that they were all extremely pessimistic (with a brief lapse into modified optimism in September, corrected in October). Nevertheless, in the light of the actual course of events they were scarcely pessimistic enough. Nor do I see any reason for expecting any appreciable alleviation in the coming months." His September 1930 letter reported that "An extraordinary example of the way in which a situation can suddenly turn round, when a tendency has been greatly overdone, has been seen on the London Stock Exchange in the last two weeks. There has been no recovery of business in Great Britain to account for it. The real facts are much as they were a month ago. But market pessimism, aided by bear operations, had brought security prices down to an absurdly low level not justified by the circumstances … everyone knew in his heart that prices were falling to foolish levels. The result was that within a few days the prices of many leading securities had risen from 10 to 20 per cent." The stock market had diverged from any level that could be construed as reflecting underlying fundamentals, but then abruptly bounced back. Keynes again stressed that Britain was not doing as badly as the United States in the slump: the fall in the British index of production from the previous year "is certainly less than 10 per cent" whereas the US index of industrial production for July 1930 was 37% below that for July 1929.

Keynes's 1930 "October Letter" warned that, "The catastrophic increase in the value of money has raised the burden of indebtedness of many countries beyond what they can bear … in many parts of the world the fall of prices has now reached a point where it is straining the social system at its foundations. Agriculturists and other producers of primary materials are being threatened with ruin and bankruptcy all over the world. It is useless to expect a recovery of markets in such conditions" (and in his February 1931 letter he again warned that "The prospect of a long series of defaults [by debtor countries exporting raw materials] during 1931 is not be excluded"). All of the gains that Germany had received in the Young Plan for reparations compared to the Dawes Plan were obliterated because "the clause in the Dawes Plan by which her [Germany's] liabilities in terms of gold were to be modified in the event of a change in prices was not included in the Young Plan." Keynes declared himself "rather more pessimistic … than a month ago." He remarked that in Britain, "Very slight steps have been taken, as yet, in the direction of reducing wages, which is probably inevitable, but will not get anyone much further if all countries alike embark on wage-cutting policies."

These themes of Keynes's October 1930 letter to Philips, the danger of ruin and bankruptcy from price deflation in a world where debts are fixed in money terms and the futility of wage-cutting, appeared publically in his December article in *The Nation and Atheneum* on "The Great Slump of 1930" (reprinted in his *Essays in Persuasion*, 1931). There Keynes (1931, 138–139) warned that, since wage and price deflation increases the real burden of debt and wage cuts reduce purchasing power, "neither the restriction of output nor the reduction of wages serves in itself to restore equilibrium" and went on to emphasize that "Moreover, even if we were to succeed eventually in reestablishing output at the lower level of money-wages appropriate to (say) the pre-war level of prices, our troubles would not be at an end. For since 1914 an immense burden of bonded debt, both national and international, has been contracted, which is fixed in terms of money. Thus every fall of prices increases the value of the money in which it is fixed. For example, if we were to settle down to the pre-war level of prices, the British National Debt would be nearly 40% greater than it was in 1924 and double what it was in 1920; … the obligations of such debtor countries as those of South America and Australia would become insupportable without a reduction of their standard of life for the benefit of their creditors; agriculturalists and householders throughout the world, who have borrowed o mortgage, would find themselves the victims of their creditors. In such a situation it must be doubtful whether the necessary adjustments could be made in time to prevent a series of bankruptcies, defaults, and repudiations which would shake the capitalist order to its foundations" (see also Dimand 2011). Here, before Fisher (1932, 1933, see Dimand 2019), was the concern with the effect of deflation on the real value of nominal deflation that reappeared in Chapter 19, "Changes in Money Wages," of *The General Theory*, where Keynes (1936, 264) warned that "if the fall of wages and prices goes far, the embarrassment of those entrepreneurs who are heavily indebted may soon reach the point of insolvency – with severely adverse effects on investment."

Contested Budgets, Trade Balance and the Banking and Exchange Crises of 1931

In 1930, Keynes's "November Letter" argued that foreign opinion underestimated the financial strength that accompanied Britain's industrial weakness: "it is forgotten that the adverse tendencies of the foreign exchanges, until recently, have been due, not to the absence of a favorable foreign trade balance, but to the eagerness of British investors to take advantage of the high profits or high rates of interest obtainable abroad. In 1929 the British favorable balance available for new foreign investment was greater than that for any other country, greater even than that for the United States. The Bank of England's difficulties were due to the fact that the pressure of savers to take advantage of opportunities abroad was even greater." Subsequent events in Wall Street and elsewhere had made overseas investment less appealing to British savers, so that the Bank of England was holding twenty million pounds sterling more of gold than a year before. In his December 1930 letter, Keynes reported that, even though "The perpetual drain of gold to France provides a source of nervousness and irritation in the money market" and although thirty million pounds sterling of gold had moved from Britain to France in the previous three months, the Bank of England held twenty-two million pounds sterling more in gold than a year before (but Keynes's March 1931 letter reported that a drain of twenty million pounds sterling of gold from the Bank of England in the previous three months "causing nervous talk to prevail in London"). Despite Keynes's repeated insistence on the financial strength of sterling and the growing gold reserves of the Bank of England (less than a year before the crisis of August and September 1931 that forced Britain off the gold exchange standard), the underlying message was that capital mobility under fixed exchange rates would constrain even the Bank of England from trying to lower long-term interest rates to stimulate investment. Until Britain left the gold standard and allowed sterling to float, Keynes's letters to Philips monitored the strength of protectionist sentiment in the British Government, but he lost interest in tariff proposals once the exchange rate was no longer pegged (see Keynes 1931). But there was one bright spot for Britain: Keynes's February 1931 letter stressed that "It must not be overlooked that England is gaining enormously by the tremendous drop in the price of her imports as compared with that of her exports."

Keynes's April 1931 letter to Philips is notable for explaining that Britain's apparent budget deficit of £23.5 million for the fiscal year ending March 31 "is not as bad as it sounds, since this figure is reached after allowing for the repayment of £67,000,000 of debt. So that, apart from debt repayments, there was a surplus on the year's workings of £43,500,000. It must be doubtful whether any other country is showing so favorable a result. Even if the sum borrowed for the unemployment fund, which lies outside the budget[5], were to be deducted, there would still have on the year a net reduction of debt." The next year's was expected to be larger, but "If no debt were to be repaid, there would probably be no deficit, even for the forthcoming year." Keynes's May 1931 letter, reporting on the budget presented by Labor Chancellor of the Exchequer Phillip Snowden, noted that "there will still be some reduction of debt during the forthcoming year, though not on as large as a scale as formerly." A few months later, when Snowden and Prime Minister Ramsay MacDonald broke with their party to join the Conservatives in a National Government to deal with a budget and exchange crisis, Snowden found it convenient to overlook that the apparent budget deficit was an artifact of budgeting for a reduction in the national debt, and to denounce his former Labor Cabinet colleagues for endangering the savings of small depositors by having the Post Office Savings Bank lend to the Unemployment Insurance Fund, without mentioned that such loans were guaranteed by the Treasury or that he had neglected to inform his Cabinet colleagues of the borrowing (as Keynes indignantly explained in two paragraphs in the draft of his November 1931 letter, deleted from the final version).

Keynes's May 1931 letter is also notable, in light of the subsequent exchange crisis that forced Britain off gold in September, for insisting that "The improvement in the sterling exchanges and the better gold position of the Bank of England, as it appears in the public returns, are not deceptive and may be assessed at even more than their face value." He held that "When there is no longer serious pressure on the Bank of England's gold, the stage will be set for really cheap money throughout the world … It will not mean a recovery, but it will pave the way for the recovery of investment which must precede the recovery of prices and profits." Keynes again emphasized that "the fall in the prices of the commodities imported by Great Britain has been so much greater than the fall in the prices of her exports. On the visible trade balance Great Britain was £5,000,000 better off in the first quarter of 1931 than in either of the preceding years … Thus the main burden of the present crisis falls on the raw-material-producing countries, and Great Britain is likely to gain gold in spite of the immense decline of her exports."

By the next month, as the Credit-Anstalt collapsed in Vienna (see Schubert 1991), as French and American capital then took flight from Germany (see Balderston 1994), and as share prices slumped in London, Wall Street and on most European bourses, Keynes felt "that we are now entering the crisis, or panic, phase of the slump. I am inclined to think that we look back on this particular slump we shall feel that this phase has been reached in the summer months of 1931, rather than at any earlier date." He warned that "the consequences of a change in the value of money, as reflected in the prices of leading commodities, so violent as that which has occurred in the last eighteen months, cannot be regarded too gravely. Until prices show a material rise the whole fabric of economic society will be shaken. Each decline of commodity prices and each further collapse on the Stock Exchanges of the world brings a further group of individuals or institutions into a position where their assets doubtfully exceed their liabilities."

Looking across the Atlantic: The American Slump

Keynes's July 1931 letter focused on the United States, where 21% of the industrial population was unemployed with perhaps another 20% working only two or three days a week: "it is quite out of the question that there should be anything which could be called a true recovery of trade at any time within, say, the next nine months. The necessary foundations for such a recover simply do not exist." Many of the loans of small banks to farmers or secured by real estate "are non-liquid and probably impaired. Thus there is a strong desire for the utmost liquidity while obtainable on the part of the ordinary Bank; and general unwillingness to take any unnecessary risks or to embark on speculative enterprise, even where the risk may be actuarially a sound one. The nervousness on the part of the Bankers is accompanied by a nervousness of the part of their depositors … So there is quite a common tendency to withdraw money from the banks and keep resources hoarded in actual cash … It was estimated that in the country as a whole as much as $500,000,000 was hoarded in actual cash in this way" (see Fisher 1933, Friedman and Schwartz 1963, Bernanke 2000). Keynes stressed that, "The American financial structure is more able than the financial structure of the European countries to support the strain of so great a change in the value of money. The very great development of Bank deposit and of bondage indebtedness in the United States means that a money contract has been interposed between the real estate on the one hand and the ultimate owner of the wealth on the other. The depreciation in the money value of the real estate sufficient to cause margins to run off, necessarily tends therefore to threaten the solidity of the structure."

Keynes reported in his July 1931 letter that although US agricultural wages had fallen by 20 to 25%, and there had also been large cuts to wages in small-scale industrial enterprises, hourly wages were practically unchanged for two thirds of the workers in large-scale industrial enterprises while the hourly wages of the other third had been reduced by some 10%. In October 1934, however, Keynes stated in his Cambridge lectures that "Labor will and has accepted reductions in money wages, in the USA in 1932, and it will not serve to reduce unemployment" with one student's notes calling the money-wage reductions "catastrophic" (Rymes 1987, 131).

Germany Defaults, Britain Abandons the Gold Parity

Turning from the United States, Keynes remarked near the end of his July letter that, "At the moment of writing there are heavy gold drains from London; but I do not think that this need be regarded with any undue alarm," a judgment that proved too sanguine.

More presciently, he added "The real danger in the situation comes from the possibility of the declaration of a general moratorium in Germany and the collapse of the mark [Germany defaulted on July 15]. The repercussion of such events on the solvency of the banking and money market systems of the world would be most serious." The next month, in his August 1931 newsletter (dated August 4), Keynes reported that "the bulk of the remaining short-term German debt is due to British and American banks and accepting houses; many accepting houses being landed with what are certainly frozen and may prove doubtful debts. Their own credit has suffered with the inevitable result, since they were the holders of large foreign balances, of a drain of gold from London … it would seem to be only ordinary prudence to act on the assumption that, while worse developments in Germany are doubtless possible, even apart from this the general underlying position is worse than the ordinary reader of newspapers believes it to be." While "Great Britain is suffering from the temporary shock to confidence due to the difficulties of the accepting houses,"[6] the situation of the world economy as a whole was more serious: "We are certainly standing in the midst of the greatest economic crisis of the modern world. Important though the German developments have been I would emphasize that these have been essentially consequences of deeper causes which are affecting all countries alike … For there is no financial structure which can withstand the strain of so violent a disturbance of values." A handwritten postscript at the end of the typed August 1931 letter warns Keynes's readers "not to be encouraged even by the appearance of apparently good news. The world financial structure is shaken and is rotten in many directions. Patching arrangements will be attempted, but they will not do much good, and it would be a mistake to place reliance on them." The next day, August 5, Keynes, writing to Prime Minister J. Ramsay MacDonald to urge rejection of the May Report, stated that "it is now virtually *certain* that we shall go off the existing parity at no distant date … when doubts, as to the prosperity of a currency, such as now exist about sterling, have come into existence, the game's up" (Keynes 1971–1989, Vol. XX, 591–593; Skidelsky 2003, 446), but he did not say so in print or to Philips – and he rejected, on patriotic grounds, a suggestion by O. T. Falk that the Independent Investment Trust, of which Keynes and Falk were directors, should replace a dollar loan with a sterling loan, which Keynes condemned as "a frank bear speculation against sterling." The Independent Investment Trust lost £40,000 by not switching its financing (Keynes 1971–1989, Vol. XX, 611–612; Moggridge 1992, 528–529; Skidelsky 2003, 447).

It was not only the world financial structure that was shaken; so was the Secretary Department of N. V. Philips. On August 6, 1931, H. du Pré wrote plaintively to Keynes, "Though we could hardly expect otherwise from your former letters, we note that you are not at all optimistic about the developments in the latter part of this year. These last weeks we read in the papers some statements from several Americans (among them people of authority), which hold a somewhat more cheerful view for the coming months. Must we infer from your letter that they are still, or again, too optimistic or is it possible that since your return from America[7] there have been some improvements, which may lead one to expect some improvement at least for the autumn?" Even Roger Babson, who had made his reputation by being bearish about the stock market in September 1929 (as he had been since 1926), was bullish by early 1931 (see W. Friedman 2014).

Keynes's reply on August 12 crushed any hopes: "In response to your enquiry, nothing has happened to make me more optimistic. As regards America, I consider that recovery this autumn is altogether out of the question. But the minds of all of us are of course dominated by the European and indeed the world situation. This still seems to me to be, as I have already described it, more serious than the general public know. I should recommend as complete inaction as is possible until further crises, or further striking events of some kind or another have occurred to clear up the situation."

Keynes's September letter (dated September 10, 1931), after the Conservative-dominated National Government displaced Labor, warned that "the hysterical concentration on Budgeting economy, which has also spread to the curtailment of expenditure by Local Authorities is calculated to produce unfavorable developments. For the widespread curtailment of expenditure is certain to reduce business profits and increase unemployment and lower the receipts of the Treasury, whilst it will do very little to tackle what is the fundamental problem, namely the improvement of the British Trade Balance. We seem likely to be faced by a period during which the balance of trade will not be sufficient to give confidence to foreign depositors."

It turned out, however, that one part of the cuts in government spending, the reduction in pay of the armed services, did indirectly dispose of the balance of payments problem. Since the government's version of equal sacrifice was that a vice-admiral earning £5 10s a day would lose 10 shillings a day (a reduction of 1/11), while naval lieutenants earning £1 7s a day and able-bodied seamen earning 5 shillings a day should each lose a shilling a day, reductions of 1/27 and 1/5, respectively (Muggeridge 1940, 109n), a naval mutiny erupted at Invergordon on September 16 (the first British naval mutiny since 1797), leading to abandonment of a fixed exchange rate on September 21 and a prompt 20% depreciation of sterling. Once the gold parity was abandoned, interest rates could be lowered without any balance of payments crisis. Commander Stephen King-Hall remarked "the strange combination of circumstances which caused the Royal Navy to be used by a far-seeing Providence as the unconscious means of … releasing the nation from the onerous terms of the contract of 1925 when the pound was restored to gold at pre-war parity … In 1805 the Navy saved the nation at Trafalgar; it may be that at Invergordon it achieved a like feat" (quoted by Muggeridge 1940, 111n). As for the budget deficit, Chancellor Snowden, who in the preceding Labor government had steadfastly blocked any reduction in the Sinking Fund contributions for paying down the national debt, now presented a budget reducing the annual Sinking Fund contribution by £20 million. Keynes declared in his October 1931 letter to Philips, "Great Britain's inevitable departure from the gold standard having occurred, it has been received with almost universal relief and in industrial circles a spirit of optimism is now abroad … Since the City and the Bank of England did their utmost to avoid the change, they feel that honor is satisfied. In other quarters the effect is to relieve a tension which was becoming almost unbearable … I have no doubt at all as to the reality of the stimulus which British business has obtained." Fisher (1935), assembling data on twenty-nine countries, found that recovery began only once a country abandoned the gold parity and was able to pursue a looser monetary policy (see Dimand 2003).

Keynes concluded his October 1931 letter, "The general passion for liquidity is bringing the value of cash in terms of everything else to so high a level as to be very near breaking point. This does not apply to Great Britain since her crisis was a balance of payments crisis rather than a banking crisis strictly so called. Thus the possibility of a general European and American banking crisis is the main risk, the possibility of which has now to be borne in mind." The US banking crisis culminated in the "Bank Holiday" of March 1933, while all the major German and Italian banks passed into government ownership.

On November 3, 1931, Dr. du Pré was "very sorry to say that the necessity for the strictest economy which makes itself felt in all departments of our concern at present, impels us to an important curtailment of the budget of our Economic Intelligence Service" which would now issue bulletins every three months, instead of monthly. He asked Keynes for quarterly letters for £50 per annum, instead of monthly letters for £150 per annum. Keynes replied on November 9 that he read the letter "without any great surprise. I had been rather hesitating in my mind as to whether it is worth while to continue the arrangement on the new basis. But on the whole I feel that I should not like to break the friendly relations which have arisen between us, merely because times are bad." He accepted the offer[8], asking to be reminded when each quarterly report was due, and enclosed his November letter stating that Britain was "to a considerable extent getting the best of both worlds since broadly speaking the countries from which we buy our food and raw materials have followed us off gold, whilst our manufacturing competitors have remained on the old gold parity."[9] He felt that Continental observers were mistaken to think that Britain would want to return to gold: "Foreigners always underestimate the slow infiltration of what I have sometimes called 'inside opinion', whilst 'outside opinion' remains ostensibly unchanged. Then quite suddenly what 'inside opinion' becomes 'outside opinion'. Foreigners are quite taken by surprise, but the change is really one which had been long prepared. In the later months of the old gold standard there was a hardly a soul in this country who really believed in it. But it was considered that it was our duty for fairly obvious reasons to do everything we possibly could to keep where we were."

Keynes's May 1932 quarterly letter stressed that, "The most important development, if one is thinking not so much of the moment but of laying the foundations for future improvement, is to be found in the return to cheap money, which was interrupted by the financial crisis of last summer and the departure from gold. I am more and more convinced in the belief, which I have held for some time, that an ultra-cheap money phase in the principal financial centers is an indispensable preliminary to recovery … Nevertheless it would be imprudent to expect too much at any early date from the stimulus of cheap money. The courage of enterprise is now so completely broken, that the effect on prices of money however cheap will be very slow. I consider it likely, therefore, that the cheap money phase may be extremely prolonged and that it may proceed to unprecedented lengths before it produces its effect." He concluded, "For the time being the world is marking time, – waiting for it does not quite know what. I emphasize again the fact that the position in Great Britain, and in some of her Dominions, is relatively good. But for the time being, I see no light anywhere else … It would certainly be much too soon to take any steps whatever to be ready for a possible revival."

Looking across the Atlantic: Hope from the New Deal

Keynes's August 1932 memorandum was notable for its explanation of why US stock prices had risen sharply and why that need not signal an end to the industrial crisis: the financial crisis had driven down stock prices until "the securities of many famous and successful companies were standing at little more than the equivalent of the net cash and liquid resources owned by those companies … the assets in question would either be worth nothing as a result of the general breakdown of contract, or must, in any circumstances apart from that, be worth a very great deal more than their quotations. Consequently, it is logical and right that the fear of their being worth nothing having been brought to an end, there should be a rapid recovery of the quotations on a very striking scale. It does not need a termination of the industrial crisis, or even an expectation of its early cessation, in order to justify the new levels."

In his February 1933 memorandum, commenting on the likely futility of the projected World Economic Conference, Keynes recalled that "I have myself put forward more drastic proposals for an international fiduciary currency, which would be the legal equivalent of gold. If this were agreed to, the position would be so much eased that various other desirable measures would also become practicable. I do not despair of converting British opinion to such a plan, but I am told that continental opinion would be almost unanimously opposed it." Keynes had contemplated such proposals long before Bretton Woods.

Keynes's August 1933 memorandum (actually mailed July 20, before Keynes left for holidays) held that "My own view is that President's Roosevelt's programme is to be taken most seriously as a means not only of American, but of world recovery. He will suffer set-backs and no one can predict the end of the story. But it does seem fairly safe to say that his drastic policies have had the result of turning the tide in the direction of better security not only in the United States, but elsewhere … Perhaps in the end President Roosevelt will devalue the dollar in terms of gold by 30 or 40 per cent." His November 1933 memorandum regretted "the failure of the President during his first six months to act inflation as well as talk it. In actual fact Governmental loan expenditure in the United States up to the end of September was on quite a trifling scale" but since then it seemed to be increasing: "if during the next six months the President is at last successful in putting into circulation a large volume of loan expenditure, I should expect a correspondingly rapid improvement in the industrial prosperity of America. This, if it occurs, would have a great influence on the rest of the world and especially on Great Britain … it might pave the way for a rate of improvement sufficiently rapid to deserve the name of real recovery." Keynes's February 1934 memorandum reported that in the United States "everything is moving strongly upwards. This is to be largely attributed to the fact that Governmental loan expenditure is now at last occurring on a large scale … the disbursement by the American Treasury of new money against borrowing has reached or is approaching \$50,000,000 weekly and should maintain this rate for a few months to come." In his August 1934 memorandum, having visited the United States since his May memorandum, he found there "a recession which is somewhat more than seasonal," aggravated since his visit by a "failure of the corn crop … so acute as to be little short of a national disaster" but the actual and prospective level of US Government loan-financed expenditure made him optimistic about prospects for the US economy in the autumn and winter. He also reported that "the view is generally held in Great Britain that the gold block countries – including Holland not less than the others – cannot permanently maintain their present parity with gold without a disaster. Now or later it seems to us certain that the necessity for devaluation will be admitted." The reports end with Keynes's November 1934 memorandum, with no correspondence in the Keynes Papers concerning the end of his relationship with the Philips firm.

Conclusion: The Message of Keynes's Reports to Philips

Keynes's letters to the Philips electronics firm reveal he perceived events in the British and world economies from the beginning of 1930 through November 1934, and provide pungent and insightful commentary. These reports high-light the importance to Keynes of cheap money as a stimulus to investment – he was not just concerned with fiscal policy as the means to recovery, however much he placed emphasis from 1933 onward on the loan-financed expenditure of the Roosevelt Administration in the US. Keynes's response to a query from du Pré is particularly interesting about Keynes's distinction between those investment expenditures that are sensitive to interest rates and those that are not. The reports stress a theme discussed more briefly in Keynes's 1931 Harris Foundation lectures in Chicago (in Wright, ed., 1931) and in Chapter 19 of *The General Theory*, and at greater length by Irving Fisher (1932, 1933) (and later by Hyman Minsky 1975): since debt are contracted in nominal terms, a rise in the purchasing power of money increases the risk of bankruptcy, repudiation and default – and it is not just actual defaults that are costly, but also the perception of increased riskiness. Keynes recognized the exceptional seriousness of the Depression, dissenting firmly from predictions of an early recovery, and he saw clearly how defending overvalued gold parities forced central banks to keep interest rates high, instead of pursuing ultra-cheap money to restore investment. This hitherto-neglected body of evidence allows one to watch the unfolding of the world economic crisis of the early 1930s through Keynes's eyes, extraordinary events as viewed and narrated by an extraordinary economist. At £12 10s per report (by no means a trivial sum at the time), N. V. Philips certainly got their money's worth.

Notes

1. "Thursday, October 24, is the first of the days which history – such as it is on the subject – identifies with the panic of 1929" (Galbraith 1961, 103–104), but already on Monday, October 21, Irving Fisher had characterized the fall in stock prices as just the "shaking out of the lunatic fringe" and on Tuesday, Charles Mitchell of the National City Bank declared that "the decline has gone too far" (Galbraith 1961, 102).
2. Philips Incandescent Lamp Works, later Philips Electronics, successor to a firm founded by Lion Philips (originally Presburg), maternal uncle of Karl Marx (Gabriel 2011, 44, 110, 291-93, 295, 299, 315, 334, 366). Although relations between uncle and nephew were "strained by politics" (Gabriel 2011, 291), Mary Gabriel (2011, 299) refers to Marx's "fund of last resort, his uncle … He had sold himself to this pragmatic businessman as a successful writer only temporarily short of cash." Gabriel (2011, 642) remarks that "Marx's dabbling in the stock market has been questioned by some scholars, who believe he may simply have wanted his uncle to believe he was engaged in 'capital' transactions, not *Capital*." After the death of Lion

Philips, his sons did not reply to Marx's letter asking for help with his daughter Laura's wedding (Gabriel 2011, 364). Anthony Sampson (1968, 95) reported that the firm's chairman Frits Philips was "a keen Moral Rearmer and a fervent anti-communist, embarrassed by the fact that his grandfather was a cousin of Karl Marx."

3. For a sense of what £150 a year might have meant to Keynes: Moggridge (1992, 508, 585) and Skidelsky (2003, 417–418, 519, 565) report that Keynes's net worth fluctuated from £44,000 at the end of 1927 to £7,815 at the end of 1929, then rising to over £506,222 at the end of 1936, dropping again to £181,244 at the end of 1938. The offer from Philips came at a particularly low point in his finances. According to Skidelsky (2003, 265) "investment, directorship and consultancy income" accounted for more than 70% of Keynes's income between 1923-24 and 1928-29 (including £1,000 a year as chairman of National Mutual Life Assurance), books and articles for another 20%, leaving no more than a tenth of income from such academic sources as teaching, examining, being secretary of the Royal Economic Society and editor of its journal, and being Bursar and a Fellow of King's College.
4. However, writing to Keynes on January 21, H. du Pré was moved "to remark that the latest figures from the Argentine which, according to the handwritten note at the bottom of your letter, you intended to enclose, were not received here, so that we cannot give you an opinion about their importance for us."
5. When the majority report of the May Committee on National Expenditure projected on July 31, 1931, that the budget deficit for 1931-32 would be £120 million, necessitating £96 million of cuts to unemployment benefits, road construction, and government and armed forces pay, it counted all borrowing by the Unemployment and Road funds as "public expenditure on current account" as well as "the usual provision for the redemption of debt" of £50 million (Winch 1969, 126–130). Keynes accused the majority on the May Committee of not "having given a moment's thought to the possible repercussions of their programme, either on the volume of unemployment or on the receipts of taxation" – he estimated it would add 250,000 to 400,000 to the unemployed, and reduce tax receipts by £70 million (*New Statesman and Nation*, August 15, 1931; Keynes 1971-89, Vol. IX, 141–145; Winch 1969, 130, Skidelsky 2003, 446).
6. With regard to Britain, Keynes noted that "There is, however, tremendous pressure of public opinion towards the Government Economy, which means in the main a reduction in the salaries of Government employees and of the allowances of the unemployed. It is equally difficult for the present [Labour] Government either to refuse or concede concessions to this trend of opinion. But if a movement in this direction takes place, which is still most doubtful, it remains exceedingly open to argument whether the result on the actual level of unemployment will be favourable."
7. Keynes had given three Harris Foundation Lectures on "An Economic Analysis of Unemployment" at the University of Chicago in June and July 1931, published in Quincy Wright, ed. (1931), and reprinted in Keynes (1971-89), Vol. XIII. These lectures mostly expounded the analysis of Keynes's *Treatise*, but the third lecture also examined the debt-deflation process, the undermining of the financial structure by an increase in the real value of debts and fall in the nominal value of collateral (Keynes 1971-89, Vol. XIII, 359–361, see Dimand 2011).
8. He also raised a "small personal matter", asking for advice on buying a new wireless set that would "have a thoroughly good loud speaker, both for voice and music reproduction and should be able to pick up distant stations such as Moscow."
9. A passage crossed-out in the draft of Keynes's November 1931 letter, in the section discussing the general election, stated that, "As has been the case in the last three or four General Elections, it is that old wretch Lord Rothermere [publisher of the *Daily Mail*] who has been dead right. It is said that he has made a profit on the crisis of £100,000, buying majorities on the Stock Exchange." Skidelsky (2003, 472) relates that Keynes "consistently lost money (his own and his friends') on the results of general elections."

References

Backhouse, Roger E., and Bradley W. Bateman, eds. 2006. *The Cambridge Companion to Keynes*. Cambridge, UK: Cambridge University Press.

Backhouse, Roger E., and Bradley W. Bateman. 2011. *Capitalist Revolutionary: John Maynard Keynes*. Cambridge, MA: Harvard University Press.

Balderston, Theo. 1994. "The Banks and the Gold Standard in the German Financial Crisis of 1931." *Financial History Review* 1 (1):43–68. https://doi.org/10.1017/S0968565000001554

Bernanke, Ben S. 2000. *Essays on the Great Depression*. Princeton, NJ: Princeton University Press.

Cargill, Thomas F. 1992. "Miscellany: Irving Fisher Comments on Benjamin Strong and the Federal Reserve in the 1930s." *Journal of Political Economy* 100 (6):1273–7. https://doi.org/10.1086/261861

Chadwyck-Healey. 1993. *The John Maynard Keynes Papers in King's College, Cambridge, 170 Reels of Microfilm*. Cambridge, UK: Chadwyck-Healey Ltd.

Cox, Jacqueline. 1993. *A Catalogue of the John Maynard Keynes Papers in King's College, Cambridge*. Cambridge, UK: Chadwyck-Healey Ltd.

Cox, Jacqueline. 1995. "Keynes: An Archivist's View." In *New Perspectives on Keynes,* Annual Supplement to History of Political Economy. Vol. 27, edited by Allin F. Cottrell and Michael S. Lawlor, 163–75. Durham, NC: Duke University Press.

Dimand, Robert W. 1988. *The Origins of the Keynesian Revolution*. Aldershot, UK: Edward Elgar Publishing.

Dimand, Robert W. 2003. "Irving Fisher on the International Transmission of Booms and Depressions through Monetary Standards." *Journal of Money, Credit and Banking* 35 (1):49–90. https://doi.org/10.1353/mcb.2003.0002

Dimand, Robert W. 2011. "The Consequences to the Banks of the Collapse of Money Values,' 1931 and 2009." In *Perspectives on Keynesian Economics*, edited by Arie Arnon, Jimmy Weinblatt, and Warren Young, 233–45. Heidelberg: Springer Verlag.

Dimand, Robert W. 2019. *Irving Fisher*. London: Palgrave Macmillan, Great Thinkers in Economics.

Dimand, Robert W., and Harald Hagemann, eds. 2019. *The Elgar Companion to John Maynard Keynes*. Cheltenham, UK.

Fisher, Irving. 1932. *Booms and Depressions*. New York: Adelphi.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1 (4):337–57. https://doi.org/10.2307/1907327

Fisher, Irving. 1935. "Are Booms and Depressions Transmitted Internationally through Monetary Standards?" *Bulletin de L'Institut Internationale de Statistique* 28 (1):1–29. Reprinted in Dimand (2003).

Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press for the National Bureau of Economic Research.

Friedman, Walter A. 2014. *Fortune Tellers: The Story of America's First Economic Forecasters*. Princeton, NJ: Princeton University Press.

Gabriel, Mary. 2011. *Love and Capital: Karl and Jenny Marx and the Birth of a Revolution*. New York: Little, Brown.

Galbraith, John Kenneth. 1961. *The Great Crash 1929, with New Introduction*. Boston, MA: Houghton Mifflin.

Harrod, Roy F. 1939. "An Essay in Dynamic Theory." *The Economic Journal* 49 (193):14–33. https://doi.org/10.2307/2225181

Keynes, John Maynard. 1983 [1923]. "Some Aspects of Commodity Markets." *Manchester Guardian Commercial, Reconstruction in Europe* 29 March 1923; reprinted in Keynes (1971–89), XII:255–66.

Keynes, John Maynard. 1983 [1923–30]. *London and Cambridge Economic Service Special Memoranda*, as reprinted in Keynes (1971–89), XII: 267–647.

Keynes, John Maynard. 1919. *The Economic Consequences of the Peace*. London: Macmillan.

Keynes, John Maynard. 1925. *The Economic Consequences of Mr. Churchill*. London: Leonard and Virginia Woolf, at the Hogarth Press.

Keynes, John Maynard. 1930. *A Treatise on Money*, 2 Volumes. London: Macmillan.

Keynes, John Maynard. 1931. *Essays in Persuasion*. London: Macmillan.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. London: Macmillan.

Keynes, John Maynard. 1971–89. *The Collected Writings of John Maynard Keynes*, 30 volumes, general editors Donald E. Moggridge and E. A. G. Robinson, volume editors Elizabeth S. Johnson and Donald E. Moggridge. London: Cambridge University Press, for the Royal Economic Society.

Minsky, Hyman P. 1975. *John Maynard Keynes*. New York: Columbia University Press.

Moggridge, Donald E. 1992. *Maynard Keynes: An Economist's Biography*. London: Routledge.

Moggridge, Donald E. 2006. "Keynes and His Correspondence." In *The Cambridge Companion to Keynes*, edited by Roger E. Backhouse and Bradley W. Bateman, 136–59. Cambridge, UK: Cambridge University Press.

Muggeridge, Malcolm. 1940. *The Thirties: 1930-40 in Great Britain*. London: Hamish Hamilton.

Rymes, Thomas K., ed. 1987. *Keynes's Lectures: Notes of Students*. Ottawa: Department of Economics, Carleton University.

Rymes, Thomas K., ed. 1989. *Keynes's Lectures, 1932–35: Notes of a Representative Student*. London: University of Michigan Press.

Sampson, Anthony. 1968. *The New Europeans*. London: Hodder and Stoughton.

Schubert, A. 1991. *The Credit-Anstalt Crisis of 1931*. Cambridge, UK: Cambridge University Press.

Skidelsky, Robert. 1983–2000. *John Maynard Keynes*, 3 volumes. London: Viking.

Skidelsky, Robert. 2003. *John Maynard Keynes, 1883-1946: Economist, Philosopher, Statesman*. London: Macmillan.

Winch, Donald. 1969. *Economics and Policy: A Historical Study*. New York: Walker and Company.

Wright, Quincy, ed. 1931. *Unemployment as a World Problem: The Harris Foundation Lectures*. Chicago, IL: University of Chicago Press.

Journal of Econometrics 234 (2023) 758–776

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Regression-adjusted estimation of quantile treatment effects under covariate-adaptive randomizations

Liang Jiang [a], Peter C.B. Phillips [b,c,d,e], Yubo Tao [f,*], Yichong Zhang [b]

[a] Fanhai International School of Finance, Fudan University 220 Handan Rd, Shanghai, 200437, China
[b] School of Economics, Singapore Management University 90 Stamford Rd, 178903, Singapore
[c] Yale University, New Haven, CT 06520-8281, United States of America
[d] University of Auckland, 12 Grafton Rd, Auckland Central, Auckland, 1010, New Zealand
[e] University of Southampton, University Rd, Southampton SO17 1BJ, United Kingdom
[f] Department of Economics, Faculty of Social Sciences, University of Macau Avenida da Universidade, Taipa, Macao Special Administrative Region of China

A R T I C L E I N F O

Article history:
Received 21 November 2021
Received in revised form 26 July 2022
Accepted 30 August 2022
Available online 30 November 2022

JEL classification:
C14
C21
D14
G21

Keywords:
Covariate-adaptive randomization
High-dimensional data
Regression adjustment
Quantile treatment effects

A B S T R A C T

Datasets from field experiments with covariate-adaptive randomizations (CARs) usually contain extra covariates in addition to the strata indicators. We propose to incorporate these additional covariates via auxiliary regressions in the estimation and inference of unconditional quantile treatment effects (QTEs) under CARs. We establish the consistency and limit distribution of the regression-adjusted QTE estimator and prove that the use of multiplier bootstrap inference is non-conservative under CARs. The auxiliary regression may be estimated parametrically, nonparametrically, or via regularization when the data are high-dimensional. Even when the auxiliary regression is misspecified, the proposed bootstrap inferential procedure still achieves the nominal rejection probability in the limit under the null. When the auxiliary regression is correctly specified, the regression-adjusted estimator achieves the minimum asymptotic variance. We also discuss forms of adjustments that can improve the efficiency of the QTE estimators. The finite sample performance of the new estimation and inferential methods is studied in simulations, and an empirical application to a well-known dataset concerned with expanding access to basic bank accounts on savings is reported.

1. Introduction

Covariate-adaptive randomizations (CARs) have recently seen growing use in a wide variety of randomized experiments in economic research. Examples include Chong et al. (2016), Greaney et al. (2016), Jakiela and Ozier (2016), Burchardi et al. (2019), Anderson and McKenzie (2021), among many others. In CAR modeling, units are first stratified using some baseline covariates, and then, within each stratum, the treatment status is assigned (independent of covariates) to achieve the balance between the numbers of treated and control units.

In many empirical studies, apart from the average treatment effect (ATE), researchers are often interested in using randomized experiments to estimate quantile treatment effects (QTEs). The QTE has a useful role as a robustness check for the ATE and characterizes any heterogeneity that may be present in the sign and magnitude of the treatment effects

* Corresponding author.
E-mail address: yubotao@um.edu.mo (Y. Tao).

https://doi.org/10.1016/j.jeconom.2022.08.010
0304-4076/

according to their position within the distribution of outcomes. See, for example, Bitler et al. (2006), Muralidharan and Sundararaman (2011), Duflo et al. (2013), Banerjee et al. (2015), Crépon et al. (2015), and Campos et al. (2017).

Two practical issues arise in estimation and inference concerning QTEs under CARs. First, other covariates in addition to the strata indicators are collected during the experiment. It is possible to incorporate these covariates in the estimation of treatment effects to reduce variance and improve efficiency. In the estimation of ATE, the usual practice is to run a simple ordinary least squares (OLS) regression of the outcome on treatment status, strata indicators, and additional covariates as in the analysis of covariance (ANCOVA). Freedman (2008a,b) pointed out that such an OLS regression adjustment can degrade the precision of the ATE estimator. Lin (2013) reexamined Freedman's critique and showed that, in order to improve efficiency, the linear regression adjustment should include a full set of interactions between the treatment status and covariates. However, because the quantile function is a nonlinear operator, even when the assignment of treatment status is completely random, a similar linear quantile regression with a full set of interaction terms is unable to provide a consistent estimate of the *unconditional* QTE, not to mention the improvement of estimation efficiency. Second, in order to achieve balance in the respective number of treated and control units within each stratum, treatment statuses under CARs usually exhibit a (negative) cross-sectional dependence. Standard inference procedures that rely on cross-sectional independence are therefore conservative and lack power. These two issues raise questions of how to use the additional covariates to consistently and more efficiently estimate QTE in CAR settings and how to conduct valid statistical procedures that mitigate conservatism in inference.[1]

The present paper addresses these issues by proposing a regression-adjusted estimator of the QTE, deriving its limit theory, and establishing the validity of multiplier bootstrap inference under CARs. Even under potential misspecification of the auxiliary regressions, the proposed QTE estimator is shown to maintain its consistency, and the multiplier bootstrap procedure is shown to have an asymptotic size equal to the nominal level under the null. When the auxiliary regression is correctly specified, the QTE estimator achieves minimum asymptotic variance.

We further investigate efficiency gains that materialize from the regression adjustments in three scenarios: (1) parametric regressions, (2) nonparametric regressions, and (3) regressions with regularization in high-dimensional settings. Specifically, for parametric regressions with a potentially misspecified linear probability model, we propose to compute the optimal linear coefficient by minimizing the variance of the QTE estimator. Such an adjustment is optimal within the class of linear adjustments but does not necessarily achieve the global minimum asymptotic variance. However, as no adjustment is a special case of the linear regression with all the coefficients being zero, our optimal linear adjustment is guaranteed to be weakly more efficient than the QTE estimator with no adjustments, which addresses Freedman's critique. We also consider a potentially misspecified logistic regression with fixed-dimensional regressors and strata- and quantile-specific regression coefficients, which is then estimated by the quasi maximum likelihood estimation (QMLE). Although the QMLE does not necessarily minimize the asymptotic variance of the QTE, such a flexible logistic model can closely approximate the true specification. Therefore, in practice, the corresponding regression-adjusted QTE estimator usually has a smaller variance than that with no adjustments. Last, we propose to treat the logistic QMLE adjustments as new linear regressors and re-construct the corresponding optimal linear adjustments. We then show the QTE estimator with the new adjustments is weakly more efficient than that with both the original logistic QMLE adjustments and no adjustments.

In nonparametric regressions, we further justify the QMLE by letting the regressors in the logistic regression be a set of sieve basis functions with increasing dimension and show how such a nonparametric regression-adjusted QTE estimator can achieve the global minimum asymptotic variance. For high-dimensional regressions with regularization, we consider logistic regression under ℓ_1 penalization, an approach that also achieves the global minimum asymptotic variance. All the limit theories hold uniformly over a compact set of quantile indices, implying that our multiplier bootstrap procedure can be used to conduct inference on QTEs involving single, multiple, or a continuum of quantile indices.

These results, including the limit distribution of the regression-adjusted QTE estimator and the validity of the multiplier bootstrap, provide novel contributions to the literature in three respects. First, the data generated under CARs are different from observational data as the observed outcomes and treatment statuses are cross-sectionally dependent due to the randomization schemes. Recently Bugni et al. (2018) established a rigorous asymptotic framework to study the ATE estimator under CARs and pointed out the conservatism of the two-sample t-test except for some special cases. (See Bugni et al. (2018, Remark 4.2) for more detail.) Our analysis follows this new framework, which departs from the literature of causal inference under an i.i.d. treatment structure.

Second, we contribute to the literature on causal inference under CARs by developing a new methodology that includes additional covariates in the estimation of the *unconditional* QTE and by establishing a general theory for regression adjustments that allow for parametric, nonparametric, and regularized estimations of the auxiliary regressions. As mentioned earlier, unlike ATE estimation, the naive linear quantile regression with additional covariates cannot even produce a consistent estimator of the QTE. Instead, we propose a new way to incorporate additional covariates based on the Neyman orthogonal moment and investigate the asymptotic properties and the efficiency gains of the proposed regression-adjusted estimator under CARs. This new machinery allows us to study the QTE regression, which is nonparametrically specified, with both linear (linear probability model) and nonlinear (logit and probit models) regression

[1] For example, Bugni et al. (2018) and Zhang and Zheng (2020) have shown that the usual two-sample t-test for inference concerning ATE and multiplier bootstrap inference concerning QTE are in general conservative under CARs.

adjustments. To clarify this contribution to the literature we note that Hu and Hu (2012), Ma et al. (2015, 2020), Olivares (2021), Shao and Yu (2013), Zhang and Zheng (2020), Ye (2018), Ye and Shao (2020) considered inference of various causal parameters under CARs but without taking into account additional covariates. Bugni et al. (2018), Bugni et al. (2019), and Bugni and Gao (2021) considered saturated regressions for ATE and local ATE, which can be viewed as regression-adjustments where strata indicators are interacted with the treatment or instrument. Shao et al. (2010) showed that if a test statistic is constructed based on the correctly specified model between outcome and additional covariates and the covariates used for CAR are functions of additional covariates, then the test statistic is valid conditional on additional covariates. Bloniarz et al. (2016), Fogarty (2018), Lin (2013), Lu (2016), Lei and Ding (2021), Li and Ding (2020), Liu et al. (2020), Liu and Yang (2020), Negi and Wooldridge (2020), Ye et al. (2022), Zhao and Ding (2021) studied various estimation methods based on regression adjustments, but these studies all focused on ATE estimation. Specifically, Liu et al. (2020) considered *linear* adjustments for *ATE* under CARs in which the covariates can be high-dimensional and the adjustments can be estimated by Lasso. Ansel et al. (2018) considered regression adjustment using additional covariates for ATE and Local ATE. We differ from them by considering *QTE* with *nonlinear* adjustments such as logistic Lasso.

Third, we establish the validity of the multiplier bootstrap inference for the regression-adjusted QTE estimator under CARs. To the best of our knowledge, Shao et al. (2010) and Zhang and Zheng (2020) are the only works in the literature that studied bootstrap inference under CARs. Shao et al. (2010) considered the covariate-adaptive bootstrap for the linear regression model. Zhang and Zheng (2020) proposed to bootstrap inverse propensity score weighted (IPW) QTE estimator with the estimated target fraction of treatment even when the truth is known. They showed that the asymptotic variance of the IPW estimator is the same under various CARs. Thus, even though the bootstrap sample ignores the cross-sectional dependence and behaves as if the randomization scheme is simple, the asymptotic variance of the bootstrap analog is still the same. We complement this research by studying the validity of multiplier bootstrap inference for our *regression-adjusted QTE* estimator. We establish analytically that the multiplier bootstrap with the estimated fraction of treatment is not conservative in the sense that it can achieve an asymptotic size equal to the nominal level under the null even when the auxiliary regressions are misspecified.

The present paper also comes under the umbrella of a growing literature that has addressed estimation and inference in randomized experiments. In this connection, we mention the studies of Hahn et al. (2011), Athey and Imbens (2017), Abadie et al. (2018), Tabord-Meehan (2021), Bai et al. (2021), Bai (2022), Jiang et al. (2021) among many others. Bai (2022) showed an 'optimal' matched-pair design can minimize the mean-squared error of the *difference-in-means* estimator for ATE, conditional on covariates. Tabord-Meehan (2021) designed an adaptive randomization procedure which can minimize the variance of the *weighted* estimator for ATE. Both works rely on a pilot experiment to design the optimal randomization. In contrast, we take the randomization scheme (i.e., CARs) as given and search for new estimators (other than *difference-in-quantile* and *weighted* estimators) for QTE that have smaller variance. In addition, our approach does not require a pilot experiment. Therefore, our and their methods are applied to different scenarios depending on the definition of 'optimality' and the data available, and thus, complement to each other.

From a practical perspective, our estimation and inferential methods have four advantages. First, they allow for common choices of auxiliary regressions such as linear probability, logit, and probit regressions, even though these regressions may be misspecified. Second, the methods can be implemented without tuning parameters. Third, our (bootstrap) estimator can be directly computed via the subgradient condition, and the auxiliary regressions need not be re-estimated in the bootstrap procedure, both of which save considerable computation time. Last, our estimation and inference methods can be implemented without the knowledge of the exact treatment assignment rule used in the experiment. This advantage is especially useful in subsample analysis, where sub-groups are defined using variables other than those to form the strata and the treatment assignment rule for each sub-group becomes unknown. See, for example, the anemic subsample analysis in Chong et al. (2016) and Zhang and Zheng (2020). These last three points are carried over from Zhang and Zheng (2020) and logically independent of the regression adjustments. One of our contributions is to show these results still hold for our regression-adjusted estimator.

The remainder of the paper is organized as follows. Section 2 describes the model setup and notation. Section 3 develops the asymptotic properties of our regression-adjusted QTE estimator. Section 4 studies the validity of the multiplier bootstrap inference. Section 5 considers parametric, nonparametric, and regularized estimation of the auxiliary regressions. Section 6 reports simulation results, and an empirical application of our methods to the impact of expanding access to basic bank accounts on savings is provided in Section 7. Section 8 concludes. Proofs of all results and some additional simulation results are given in the Online Supplement.

2. Setup and notation

Potential outcomes for treated and control groups are denoted by $Y(1)$ and $Y(0)$, respectively. Treatment status is denoted by A, with $A = 1$ indicating treated and $A = 0$ untreated. The stratum indicator is denoted by S, based on which the researcher implements the covariate-adaptive randomization. The support of S is denoted by $\mathcal{S}$, a finite set. After randomization, the researcher can observe the data $\{Y_i, S_i, A_i, X_i\}_{i\in[n]}$ where $[n] = \{1, 2, \ldots, n\}$, $Y_i = Y_i(1)A_i + Y_i(0)(1-A_i)$ is the observed outcome, and X_i contains extra covariates besides S_i in the dataset. The support of X is denoted Supp(X). In this paper, we allow X_i and S_i to be dependent. For $i \in [n]$, let $p(s) = \mathbb{P}(S_i = s)$, $n(s) = \sum_{i\in[n]} 1\{S_i = s\}$, $n_1(s) = \sum_{i\in[n]} A_i 1\{S_i = s\}$, and $n_0(s) = n(s) - n_1(s)$. We make the following assumptions on the data generating process (DGP) and the treatment assignment rule.

Assumption 1.

(i) $\{Y_i(1), Y_i(0), S_i, X_i\}_{i\in[n]}$ is i.i.d.
(ii) $\{Y_i(1), Y_i(0), X_i\}_{i\in[n]} \perp\!\!\!\perp \{A_i\}_{i\in[n]}|\{S_i\}_{i\in[n]}$.
(iii) Suppose $p(s)$ is fixed with respect to (w.r.t.) n and is positive for every $s \in \mathcal{S}$.
(iv) Let $\pi(s)$ denote the target fraction of treatment for stratum s. Then, $c < \min_{s\in\mathcal{S}} \pi(s) \leq \max_{s\in\mathcal{S}} \pi(s) < 1 - c$ for some constant $c \in (0, 0.5)$ and $\frac{D_n(s)}{n(s)} = o_p(1)$ for $s \in \mathcal{S}$, where $D_n(s) = \sum_{i\in[n]}(A_i - \pi(s))1\{S_i = s\}$.

Several remarks are in order. First, Assumption 1(i) allows for cross-sectional dependence among treatment statuses ($\{A_i\}_{i\in[n]}$), thereby accommodating many covariate-adaptive randomization schemes as discussed below. Second, although treatment statuses are cross-sectionally dependent, they are independent of the potential outcomes and additional covariates conditional on the stratum indicator S. Therefore, data are still experimental rather than observational. Third, Assumption 1(iii) requires the size of each stratum to be proportional to the sample size. Fourth, we can view $\pi(s)$ as the target fraction of treated units in stratum s. Similar to Bugni et al. (2019), we allow the target fractions to differ across strata. Just as for the overlapping support condition in an observational study, the target fractions are assumed to be bounded away from zero and one. In randomized experiments, this condition usually holds because investigators can determine $\pi(s)$ in the design stage. In fact, in most CARs, $\pi(s) = 0.5$ for $s \in \mathcal{S}$. Fifth, $D_n(s)$ represents the degree of imbalance between the real and target factions of treated units in the sth stratum. Bugni et al. (2018) show that Assumption 1(iv) holds under several covariate-adaptive treatment assignment rules such as simple random sampling (SRS), biased-coin design (BCD), adaptive biased-coin design (WEI), and stratified block randomization (SBR). For completeness, we briefly repeat their descriptions below. Note we only require $D_n(s)/n(s) = o_p(1)$, which is weaker than the assumption imposed by Bugni et al. (2018) but the same as that imposed by Bugni et al. (2019) and Zhang and Zheng (2020).

Example 1 (*SRS*)**.** Let $\{A_i\}_{i\in[n]}$ be drawn independently across i and of $\{S_i\}_{i\in[n]}$ as Bernoulli random variables with success rate π, i.e., for $k = 1, \ldots, n$,

$$\mathbb{P}\left(A_k = 1 \middle| \{S_i\}_{i\in[n]}, \{A_j\}_{j\in[k-1]}\right) = \mathbb{P}(A_k = 1) = \pi(S_i).$$

Example 2 (*WEI*)**.** This design was first proposed by Wei (1978). Let $n_{k-1}(S_k) = \sum_{i\in[k-1]} 1\{S_i = S_k\}$, $D_{k-1}(s) = \sum_{i\in[k-1]} \left(A_i - \frac{1}{2}\right) 1\{S_i = s\}$, and

$$\mathbb{P}\left(A_k = 1 \middle| \{S_i\}_{i\in[k]}, \{A_i\}_{i\in[k-1]}\right) = \phi\left(\frac{2D_{k-1}(S_k)}{n_{k-1}(S_k)}\right),$$

where $\phi(\cdot) : [-1, 1] \mapsto [0, 1]$ is a pre-specified non-increasing function satisfying $\phi(-x) = 1 - \phi(x)$ and $\frac{D_0(S_1)}{0}$ is understood to be zero.

Example 3 (*BCD*)**.** The treatment status is determined sequentially for $1 \leq k \leq n$ as

$$\mathbb{P}\left(A_k = 1|\{S_i\}_{i\in[k]}, \{A_i\}_{i\in[k-1]}\right) = \begin{cases} \frac{1}{2} & \text{if } D_{k-1}(S_k) = 0 \\ \lambda & \text{if } D_{k-1}(S_k) < 0 \\ 1 - \lambda & \text{if } D_{k-1}(S_k) > 0, \end{cases}$$

where $D_{k-1}(s)$ is defined as above and $\frac{1}{2} < \lambda \leq 1$.

Example 4 (*SBR*)**.** For each stratum, $\lfloor \pi(s)n(s) \rfloor$ units are assigned to treatment and the rest are assigned to control.

Denote the τth quantile of $Y(a)$ by $q_a(\tau)$ for $a = 0, 1$. We are interested in estimating and inferring the τth quantile treatment effect defined as $q(\tau) = q_1(\tau) - q_0(\tau)$. The testing problems of interest involve single, multiple, or even a continuum of quantile indices, as in the following null hypotheses

$$\begin{aligned} &\mathcal{H}_0 : q(\tau) = \underline{q} \quad \text{versus} \quad q(\tau) \neq \underline{q}, \\ &\mathcal{H}_0 : q(\tau_1) - q(\tau_2) = \underline{q} \quad \text{versus} \quad q(\tau_1) - q(\tau_2) \neq \underline{q}, \text{ and} \\ &\mathcal{H}_0 : q(\tau) = \underline{q}(\tau)\ \forall \tau \in \Upsilon \quad \text{versus} \quad q(\tau) \neq \underline{q}(\tau) \text{ for some } \tau \in \Upsilon, \end{aligned}$$

for some pre-specified value $\underline{q}$ or function $\underline{q}(\tau)$, where Υ is some compact subset of $(0, 1)$. We can also test constant QTE by letting $\underline{q}(\tau)$ in the last hypothesis be a constant $\underline{q}$.

3. Estimation

Define $m_a(\tau, s, x) = \tau - \mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x)$ for $a = 0, 1$ which are the true specifications but unknown to researchers. Instead, researchers specify working models $\{\overline{m}_a(\tau, s, x)\}_{a=0,1}$[2] for the true specification, which can be misspecified. Last, the researchers estimate the (potentially misspecified) working models via some forms of regression, and the estimators are denoted as $\{\widehat{m}_a(\tau, s, x)\}_{a=0,1}$. We also refer to $\overline{m}_a(\cdot)$ as the auxiliary regression.

Our regression-adjusted estimator of $q_1(\tau)$, denoted as $\hat{q}_1^{adj}(\tau)$, can be defined as

$$\hat{q}_1^{adj}(\tau) = \arg\min_q \sum_{i\in[n]} \left[\frac{A_i}{\hat{\pi}(S_i)} \rho_\tau(Y_i - q) + \frac{(A_i - \hat{\pi}(S_i))}{\hat{\pi}(S_i)} \widehat{m}_1(\tau, S_i, X_i) q \right], \tag{3.1}$$

where $\rho_\tau(u) = u(\tau - 1\{u \leq 0\})$ is the usual check function and $\hat{\pi}(s) = n_1(s)/n(s)$. We emphasize that $\widehat{m}_1(\cdot)$ may not consistently estimate the true specification $m_1(\cdot)$. Similarly, we can define

$$\hat{q}_0^{adj}(\tau) = \arg\min_q \sum_{i\in[n]} \left[\frac{1 - A_i}{1 - \hat{\pi}(S_i)} \rho_\tau(Y_i - q) - \frac{(A_i - \hat{\pi}(S_i))}{1 - \hat{\pi}(S_i)} \widehat{m}_0(\tau, S_i, X_i) q \right]. \tag{3.2}$$

Then, our regression adjusted QTE estimator is

$$\hat{q}^{adj}(\tau) = \hat{q}_1^{adj}(\tau) - \hat{q}_0^{adj}(\tau). \tag{3.3}$$

Several remarks are in order. First, in observational studies with i.i.d. data and $A_i \perp\!\!\!\perp X_i|S_i$, Firpo (2007), Belloni et al. (2017), and Kallus et al. (2020) showed that the doubly robust moment for $q_1(\tau)$ is

$$\mathbb{E}\left[\frac{A_i(\tau - 1\{Y_i(1) \leq q\})}{\overline{\pi}(S_i)} - \frac{A_i - \overline{\pi}(S_i)}{\overline{\pi}(S_i)} \overline{m}_1(\tau, S_i, X_i) \right] = 0, \tag{3.4}$$

where $\overline{\pi}(s)$ and $\overline{m}_1(\tau, s, x)$ are the working models for the target fraction ($\pi(s)$) and conditional probability ($m_1(\tau, s, x)$), respectively. Our estimator is motivated by this doubly robust moment, but our analysis differs from that for the observational data as CARs introduces cross-sectional dependence among observations. Second, as our target fraction estimator $\hat{\pi}(s) = n_1(s)/n(s)$ is consistent, it means $\overline{\pi}(s)$ is correctly specified as $\pi(s)$. Then, due to the double robustness, our regression adjusted estimator is consistent even when $\overline{m}_a(\cdot)$ is misspecified and $\widehat{m}_a(\cdot)$ is an inconsistent estimator of $m_a(\cdot)$. Third, we use the estimated target fraction $\hat{\pi}(s)$ even when $\pi(s)$ is known because this guarantees that the bootstrap inference is not conservative. Further discussion is provided after Theorem 4.1.

Assumption 2. For $a = 0, 1$, denote $f_a(\cdot)$, $f_a(\cdot|s)$, and $f_a(\cdot|x, s)$ as the PDFs of $Y_i(a)$, $Y_i(a)|S_i = s$, and $Y_i(a)|S_i = s, X_i = x$, respectively.

(i) $f_a(q_a(\tau))$ and $f_a(q_a(\tau)|s)$ are bounded and bounded away from zero uniformly over $\tau \in \Upsilon$ and $s \in \mathcal{S}$, where Υ is a compact subset of (0, 1).
(ii) $f_a(\cdot)$ and $f_a(\cdot|s)$ are Lipschitz over $\{q_j(\tau) : \tau \in \Upsilon\}$.
(iii) $\sup_{y\in\Re, x\in\text{Supp}(X), s\in\mathcal{S}} f_a(y|x, s) < \infty$.

Assumption 3.

(i) For $a = 0, 1$, there exists a function $\overline{m}_a(\tau, s, x)$ such that for $\overline{\Delta}_a(\tau, s, X_i) = \widehat{m}_a(\tau, s, X_i) - \overline{m}_a(\tau, s, X_i)$, we have

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i\in I_1(s)} \overline{\Delta}_a(\tau, s, X_i)}{n_1(s)} - \frac{\sum_{i\in I_0(s)} \overline{\Delta}_a(\tau, s, X_i)}{n_0(s)} \right| = o_p(n^{-1/2}),$$

where $I_a(s) = \{i \in [n] : A_i = a, S_i = s\}$.
(ii) For $a = 0, 1$, let $\mathcal{F}_a = \{\overline{m}_a(\tau, s, x) : \tau \in \Upsilon\}$ with an envelope $F_a(s, x)$. Then, $\max_{s\in\mathcal{S}} \mathbb{E}(|F_a(S_i, X_i)|^q|S_i = s) < \infty$ for $q > 2$ and there exist fixed constants $(\alpha, v) > 0$ such that

$$\sup_Q N(\mathcal{F}_a, e_Q, \varepsilon\|F_a\|_{Q,2}) \leq \left(\frac{\alpha}{\varepsilon}\right)^v, \quad \forall \varepsilon \in (0, 1],$$

where $N(\cdot)$ denotes the covering number, $e_Q(f, g) = \|f - g\|_{Q,2}$, and the supremum is taken over all finitely discrete probability measures Q.
(iii) For $a = 0, 1$ and any $\tau_1, \tau_2 \in \Upsilon$, there exists a constant $C > 0$ such that

$$\mathbb{E}((\overline{m}_a(\tau_2, S_i, X_i) - \overline{m}_a(\tau_1, S_i, X_i))^2|S_i = s) \leq C|\tau_2 - \tau_1|.$$

[2] We view $\overline{m}_a(\cdot)$ as some function with inputs τ, s, x. For example, researchers can specify a linear probability model with $\overline{m}_a(\tau, s, x) = \tau - x^\top\beta_{a,s}$, where $\beta_{a,s}$ is the linear coefficient that varies across treatment status a and stratum s.

Several remarks are in order. First, Assumption 2 is standard in the quantile regression literature. We do not need $f_a(y|x,s)$ to be bounded away from zero because we are interested in the unconditional quantile $q_a(\tau)$, which is uniquely defined as long as the unconditional density $f_a(q_a(\tau))$ is positive. Second, Assumption 3(i) is high-level. If we consider a linear probability model such that $\overline{m}_a(\tau,s,X_i) = \tau - X_i^\top \theta_{a,s}(\tau)$ and $\widehat{m}_a(\tau,s,X_i) = \tau - X_i^\top \hat{\theta}_{a,s}(\tau)$, then Assumption 3(i) is equivalent to

$$\sup_{\tau\in\Upsilon,a=0,1,s\in\mathcal{S}} \left| \left(\frac{\sum_{i\in I_1(s)} X_i}{n_1(s)} - \frac{\sum_{i\in I_0(s)} X_i}{n_0(s)} \right)^\top \left(\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau) \right) \right| = o_p(n^{-1/2}),$$

which is similar to Liu et al. (2020, Assumption 3) and holds intuitively if $\hat{\theta}_{a,s}(\tau)$ is a consistent estimator of the pseudo true value $\theta_{a,s}(\tau)$. Third, Assumptions 3(ii) and 3(iii) impose mild regularity conditions on $\overline{m}_a(\cdot)$. Assumption 3(ii) holds automatically if Υ is a finite set. In general, both Assumption 3(ii) and 3(iii) hold if

$$\sup_{a=0,1,s\in\mathcal{S},x\in\mathrm{Supp}(X)} |\overline{m}_a(\tau_2,s,x) - \overline{m}_a(\tau_1,s,x)| \le L|\tau_2-\tau_1|$$

for some constant $L > 0$. Such Lipschitz continuity holds for the true specification ($\overline{m}_a(\cdot) = m_a(\cdot)$) under Assumption 2. Fourth, we provide primitive sufficient conditions for Assumption 3 in Section 5.

Theorem 3.1. *Suppose Assumptions 1–3 hold. Then, uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}(\hat{q}^{adj}(\tau) - q(\tau)) \rightsquigarrow \mathcal{B}(\tau),$$

where $\mathcal{B}(\tau)$ is a tight Gaussian process with covariance kernel $\Sigma(\tau,\tau')$ defined in Section E of the Online Supplement. In addition, for any finite set of quantile indices $(\tau_1,\ldots,\tau_K)$, the asymptotic covariance matrix of $(\hat{q}^{adj}(\tau_1),\ldots,\hat{q}^{adj}(\tau_K))$ is denoted as $[\Sigma(\tau_k,\tau_l)]_{k,l\in[K]}$, where we use $[U_{kl}]_{k,l\in[K]}$ to denote a $K\times K$ matrix whose (k,l)th entry is U_{kl}. Then, $[\Sigma(\tau_k,\tau_l)]_{k,l\in[K]}$ is minimized in the matrix sense[3] when the auxiliary regressions are correctly specified at $(\tau_1,\ldots,\tau_K)$, i.e., $\overline{m}_a(\tau_k,s,x) = m_a(\tau_k,s,x)$ for $a = 0, 1$, $k\in[K]$, and all (s,x) in the joint support of (S_i,X_i).

Three remarks are in order. First, the expression for the asymptotic variance of $\hat{q}^{adj}(\tau)$ can be found in the proof of Theorem 3.1. It is the same whether the randomization scheme achieves strong balance[4] or not. This robustness is due to the use of the estimated target fraction ($\hat{\pi}(s)$). The same phenomenon was discovered in the simplified setting by Zhang and Zheng (2020). Second, although our estimator is still consistent and asymptotically normal when the auxiliary regression is misspecified, it is meaningful to pursue the correct specification as it achieves the minimum variance. As the estimator with no adjustments can be viewed as a special case of our estimator with $\overline{m}_a(\cdot) = 0$, Theorem 3.1 implies that the adjusted estimator with the correctly specified auxiliary regression is more efficient than that with no adjustments. If the auxiliary regression is misspecified, the adjusted estimator can sometimes be less efficient than the unadjusted one, which is known as the Freedman's critique. In Section 5, we discuss how to make adjustments that do not harm the precision of the QTE estimator. Third, the asymptotic variance of $\hat{q}^{adj}(\tau)$ depends on $(f_a(q_a(\tau)), m_a(\tau,s,x))_{a=0,1}$, which are infinite-dimensional nuisance parameters. To conduct analytic inference, it is necessary to nonparametrically estimate these nuisance parameters, which requires tuning parameters. Nonparametric estimation can be sensitive to the choice of tuning parameters and rule-of-thumb tuning parameter selection may not be appropriate for every DGP or every quantile. The use of cross-validation in selecting the tuning parameters is possible in principle but, in practice, time-consuming. These practical difficulties of analytic methods of inference provide strong motivation to investigate bootstrap inference procedures that are much less reliant on tuning parameters.

4. Multiplier bootstrap inference

We approximate the asymptotic distributions of $\hat{q}^{adj}(\tau)$ via the multiplier bootstrap. Let $\{\xi_i\}_{i\in[n]}$ be a sequence of bootstrap weights which will be specified later. Define $n_1^w(s) = \sum_{i\in[n]} \xi_i A_i 1\{S_i = s\}$, $n_0^w(s) = \sum_{i\in[n]} \xi_i(1-A_i)1\{S_i = s\}$, $n^w(s) = \sum_{i\in[n]} \xi_i 1\{S_i = s\} = n_1^w(s) + n_0^w(s)$, and $\hat{\pi}^w(s) = n_1^w(s)/n^w(s)$. The multiplier bootstrap counterpart of $\hat{q}^{adj}(\tau)$ is denoted by $\hat{q}^w(\tau)$ and defined as

$$\hat{q}^w(\tau) = \hat{q}_1^w(\tau) - \hat{q}_0^w(\tau),$$

where

$$\hat{q}_1^w(\tau) = \arg\min_q \sum_{i\in[n]} \xi_i \left[\frac{A_i}{\hat{\pi}^w(S_i)} \rho_\tau(Y_i - q) + \frac{(A_i - \hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)} \widehat{m}_1(\tau,S_i,X_i) q \right], \tag{4.1}$$

[3] For two symmetric matrices A and B, we say A is greater than or equal to B if $A - B$ is positive semidefinite.

[4] We refer readers to Bugni et al. (2018) for the definition of strong balance.

and

$$\hat{q}_0^w(\tau) = \arg\min_q \sum_{i\in[n]} \xi_i \left[\frac{1-A_i}{1-\hat{\pi}^w(S_i)} \rho_\tau(Y_i - q) - \frac{(A_i - \hat{\pi}^w(S_i))}{1-\hat{\pi}^w(S_i)} \widehat{m}_0(\tau, S_i, X_i) q \right]. \tag{4.2}$$

Two comments on implementation are noted here: (i) we do not re-estimate $\widehat{m}_a(\cdot)$ in the bootstrap sample, which is similar to the multiplier bootstrap procedure proposed by Belloni et al. (2017); and (ii) in Section B of the Online Supplement we propose a way to directly compute $(\hat{q}_a^w(\tau))_{a=0,1}$ from the subgradient conditions of (4.1) and (4.2), thereby avoiding the numerical optimization. Both features considerably reduce computation time of our bootstrap procedure.

Next, we specify the bootstrap weights.

Assumption 4. Suppose $\{\xi_i\}_{i\in[n]}$ is a sequence of nonnegative i.i.d. random variables with unit expectation and variance and a sub-exponential upper tail.

Assumption 5. Recall $\overline{\Delta}_a(\tau, s, x)$ defined in Assumption 3. We have, for $a = 0, 1$,

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i\in I_1(s)} \xi_i \overline{\Delta}_a(\tau, s, X_i)}{n_1^w(s)} - \frac{\sum_{i\in I_0(s)} \xi_i \overline{\Delta}_a(\tau, s, X_i)}{n_0^w(s)} \right| = o_p(n^{-1/2}).$$

We require the bootstrap weights to be nonnegative so that the objective functions in (4.1) and (4.2) are convex. In practice, we generate ξ_i independently from the standard exponential distribution. Assumption 5 is the bootstrap counterpart of Assumption 3. Continuing with the linear model example considered after Assumption 3, Assumption 5 requires

$$\sup_{\tau\in\Upsilon, a=0,1, s\in\mathcal{S}} \left| \left(\frac{\sum_{i\in I_1(s)} \xi_i X_i}{n_1^w(s)} - \frac{\sum_{i\in I_0(s)} \xi_i X_i}{n_0^w(s)} \right)^\top \left(\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau) \right) \right| = o_p(n^{-1/2}),$$

which holds if $\hat{\theta}_{a,s}(\tau)$ is a uniformly consistent estimator of $\theta_{a,s}(\tau)$.

Theorem 4.1. *Suppose Assumptions 1–5 hold. Then, uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}(\hat{q}^w(\tau) - \hat{q}^{adj}(\tau)) \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathcal{B}(\tau),$$

where $\mathcal{B}(\tau)$ is the same Gaussian process defined in Theorem 3.1.[5]

Two remarks are in order. First, Theorem 4.1 shows the limit distribution of the bootstrap estimator conditional on data can approximate that of the original estimator uniformly over $\tau \in \Upsilon$. This is the theoretical foundation for the bootstrap confidence intervals and bands described in Section B in the Online Supplement. Specifically, denote $\{\hat{q}^{w,b}(\tau)\}_{b\in[B]}$ as the bootstrap estimates where B is the number of bootstrap replications. Let $\widehat{\mathcal{C}}(\nu)$ and $\mathcal{C}(\nu)$ be the νth empirical quantile of the sequence $\{\hat{q}^{w,b}(\tau)\}_{b\in[B]}$ and the νth standard normal critical value, respectively. Then, we suggest using the bootstrap estimator to construct the standard error of $\hat{q}^{adj}(\tau)$ as $\hat{\sigma} = \frac{\widehat{\mathcal{C}}(0.975) - \widehat{\mathcal{C}}(0.025)}{\mathcal{C}(0.975) - \mathcal{C}(0.025)}$. Note that, unlike Hahn and Liao (2021), our bootstrap standard error is not conservative. In our context, the bootstrap estimator of σ^2 considered by Hahn and Liao (2021) is $\mathbb{E}^*(\sqrt{n}(\hat{q}^w(\tau) - \hat{q}^{adj}(\tau))^2)$, where $\mathbb{E}^*$ is the conditional expectation given data. It is well-known that weak convergence does not imply convergence in L_2-norm, which explains why they can show their estimator is in general conservative. Instead, we use a different estimator of the standard error and can show it is consistent given weak convergence. Second, such a bootstrap approximation is consistent under CAR. Zhang and Zheng (2020) showed that for the QTE estimation without regression adjustment, bootstrapping the IPW QTE estimator with the estimated target fraction results in non-conservative inference, while bootstrapping the IPW estimator with the true fraction is conservative under CARs. As the estimator considered by Zhang and Zheng (2020) is a special case of our regression-adjusted estimator with $\widehat{m}_a(\cdot) = 0$, we conjecture that the same conclusion holds. A proof of conservative bootstrap inference with the true target fraction is not included in the paper due to the space limit.[6] Our simulations confirm both the correct size coverage

[5] We view $\sqrt{n}(\hat{q}^w(\tau) - \hat{q}^{adj}(\tau))$ and $\mathcal{B}(\tau)$ as two processes indexed by $\tau \in \Upsilon$ and denote them as G_n and G, respectively. Then, following van der Vaart and Wellner (1996, Chapter 2.9), we say $G_n \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} G$ uniformly over $\tau \in \Upsilon$ if

$$\sup_{h\in \mathrm{BL}_1} |\mathbb{E}_\xi h(G_n) - \mathbb{E}h(G)| \xrightarrow{p} 0,$$

where BL_1 is the set of all functions $h : \ell^\infty(\Upsilon) \mapsto [0, 1]$ such that $|h(z_1) - h(z_2)| \le |z_1 - z_2|$ for every $z_1, z_2 \in \ell^\infty(\Upsilon)$, and $\mathbb{E}_\xi$ denotes expectation with respect to the bootstrap weights $\{\xi\}_{i\in[n]}$.

[6] Full statements and proofs are lengthy because we need to derive the limit distributions of not only the bootstrap but also the original estimator with the true target fraction. Although the negative result is theoretically interesting, we are not aware of any empirical papers using the true target fraction while making regression adjustments. Moreover, our method is shown to have better performance than the one with the true target fraction in simulations. Therefore, the practical value of proving the negative result is limited.

of our inference method using the bootstrap with the estimated target fraction and the conservatism of the bootstrap with the true target fraction. The standard error of the QTE estimator is found to be 34.9% larger on average by using the true rather than the estimated target fraction in the simulations (see Tables 1 below and A9 in the Online Supplement).

5. Auxiliary regressions

In this section, we consider two approaches to estimate the auxiliary regressions: (1) a parametric method and (2) a nonparametric method. In Section A of the Online Supplement, we further consider a regularization method. For the parametric method, we do not require the model to be correctly specified. We propose ways to estimate the pseudo true value of the auxiliary regression. For the other two methods, we (nonparametrically) estimate the true model so that the asymptotic variance of $\hat{q}^{adj}(\tau)$ achieves its minimum based on Theorem 3.1. For all three methods, we verify Assumptions 3 and 5.

5.1. Parametric method

In this section, we consider the case where X_i is finite-dimensional. Recall $m_a(\tau, s, x) \equiv \tau - \mathbb{P}(Y_i(a) \le q_a(\tau)|X_i = x, S_i = s)$ for $a = 0, 1$. We propose to model $\mathbb{P}(Y_i(a) \le q_a(\tau)|X_i, S_i = s)$ as $\Lambda_{\tau,s}(X_i, \theta_{a,s}(\tau))$, where $\theta_{a,s}(\tau)$ is a finite dimensional parameter that depends on (a, s, τ) so that our model for $m_a(\tau, s, X_i)$ is

$$\overline{m}_a(\tau, s, X_i) = \tau - \Lambda_{\tau,s}(X_i, \theta_{a,s}(\tau)). \tag{5.1}$$

We note that, as we allow for misspecification, the researchers have the freedom to choose any functional forms for $\Lambda_{\tau,s}(\cdot)$ and any pseudo true values for $\theta_{a,s}(\tau)$, both of which can vary with respect to (τ, s). For example, if we assume a logistic regression with $\Lambda_{\tau,s}(X_i, \theta_{a,s}(\tau)) = \lambda(X_i^\top \theta_{a,s}(\tau))$, where $\lambda(\cdot)$ is the logistic CDF, then there are various choices of $\theta_{a,s}(\tau)$ such as the maximizer of the population pseudo likelihood, the maximizer of the population version of the least squares objective function, or the minimizer of the asymptotic variance of the adjusted QTE estimator. As the logistic model is potentially misspecified, these three pseudo true values are not necessarily the same and can lead to different adjustments, and thus, different asymptotic variances of the corresponding adjusted QTE estimators.

Next, we first state a general result for generic choices of $\Lambda_{\tau,s}(\cdot)$ and $\theta_{a,s}(\tau)$. Suppose we estimate $\theta_{a,s}(\tau)$ by $\hat{\theta}_{a,s}(\tau)$. Then, the corresponding $\widehat{m}_a(\tau, s, X_i)$ can be written as

$$\widehat{m}_a(\tau, s, X_i) = \tau - \Lambda_{\tau,s}(X_i, \hat{\theta}_{a,s}(\tau)). \tag{5.2}$$

Assumption 6.

(i) Suppose there exist a positive random variable L_i and a positive constant $C > 0$ such that

$$\sup_{\tau_1,\tau_2\in\Upsilon, s\in\mathcal{S}, \|\theta\|\le C} \|\Lambda_{\tau_1,s}(X_i, \theta) - \Lambda_{\tau_2,s}(X_i, \theta)\|_2 \le L_i|\tau_1 - \tau_2|, \qquad \sup_{\tau\in\Upsilon, s\in\mathcal{S}, \|\theta\|\le C} |\Lambda_{\tau,s}(X_i, \theta)| \le L_i,$$
$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \|\theta\|\le C} |\partial_\theta \Lambda_{\tau,s}(X_i, \theta)| \le L_i, \quad \text{and} \quad \mathbb{E}(L_i^d|S_i = s) \le C < \infty \quad \text{for some } d > 2.$$

(ii) $\sup_{\tau_1,\tau_2\in\Upsilon, a=0,1, s\in\mathcal{S}} |\theta_{a,s}(\tau_1) - \theta_{a,s}(\tau_2)| \le C|\tau_1 - \tau_2|$.

(iii) $\sup_{\tau\in\Upsilon, a=0,1, s\in\mathcal{S}} \|\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau)\|_2 \xrightarrow{p} 0$.

Three remarks are in order. First, common choices for auxiliary regressions are linear probability, logistic, and probit regressions, corresponding to $\Lambda_{\tau,s}(X_i, \theta_{a,s}(\tau)) = X_i^\top \theta_{a,s}(\tau)$, $\lambda(\vec{X}_i^\top \theta_{a,s}(\tau))$, and $\Phi(\vec{X}_i^\top \theta_{a,s}(\tau))$, respectively, where $\Phi(\cdot)$ is the standard normal CDF and $\vec{X}_i = (1, X_i^\top)^\top$. For these models, the functional form $\Lambda_{\tau,s}(\cdot)$ does not depend on (τ, s), and Assumption 6(i) holds automatically. For the linear regression case, we do not include the intercept because our regression adjusted estimators ((3.1) and (3.2)) and their bootstrap counterparts ((4.1) and (4.2)) are numerically invariant to location shift of the auxiliary regressions. Second, it is also important to allow the functional form $\Lambda_{\tau,s}(\cdot)$ to vary across τ to incorporate the case in which the regressor X_i in the linear, logistic, and probit regressions is replaced by $W_{i,s}(\tau)$, a function of X_i that depends on (τ, s). We give a concrete example for this situation in Section 5.1.3. Third, Assumption 6(ii) also holds automatically if Υ is finite. When Υ is infinite, this condition is still mild.

Theorem 5.1. *Denote $\hat{q}^{par}(\tau)$ and $\hat{q}^{par,w}(\tau)$ as the τth QTE estimator and its multiplier bootstrap counterpart defined in Sections 3 and 4, respectively, with $\overline{m}_a(\tau, S_i, X_i)$ and $\widehat{m}_a(\tau, S_i, X_i)$ defined in (5.1) and (5.2), respectively. Suppose Assumption 1, 2, 4, and 6 hold. Then, Assumptions 3 and 5 hold, which further implies Theorems 3.1 and 4.1 hold for $\hat{q}^{par}(\tau)$ and $\hat{q}^{par,w}(\tau)$, respectively.*

Theorem 5.1 shows that, as long as the estimator of the pseudo true value $(\hat{\theta}_{a,s}(\tau))$ is uniformly consistent, under mild regularity conditions, all the general estimation and bootstrap inference results established in Sections 3 and 4 hold.

5.1.1. Linear probability model

In this section, we consider linear adjustment with parameter $t_{a,s}(\tau)$ such that

$$\Lambda_{\tau,s}(X_i, t_{a,s}(\tau)) = W_{i,s}^\top(\tau)t_{a,s}(\tau) \quad \text{and} \quad \overline{m}_a(\tau, s, X_i) = \tau - W_{i,s}^\top(\tau)t_{a,s}(\tau), \tag{5.3}$$

where the regressor $W_{i,s}(\tau)$ is a function of X_i but the functional form may vary across s, τ. For example, we can consider $W_{i,s}(\tau) = X_i$, the transformations of X_i such as quadratic and interaction terms, and some prediction of $(1\{Y_i(1) \leq q_1(\tau)\}, 1\{Y_i(0) \leq q_0(\tau)\})$ given X_i and $S_i = s$. The last example is further explained in Section 5.1.3.

We note that the asymptotic variance (denoted as σ^2) of the $\hat{q}^{adj}(\tau)$ is a function of the working model $(\overline{m}_a(\tau, s, \cdot))$, which is further indexed by its parameters (denoted as $\{t_{a,s}(\tau)\}_{a=0,1,s\in\mathcal{S}}$), i.e., $\sigma^2 = \sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}})$. Our optimal linear adjustment corresponds to parameter value $\theta_{a,s}(\tau)$ such that it minimizes $\sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}})$, i.e.,

$$\{\theta_{a,s}(\tau)\}_{a=0,1,s\in\mathcal{S}} \in \mathop{\arg\min}_{t_{a,s}:a=0,1,s\in\mathcal{S}} \sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}}).$$

Assumption 7. Define $\tilde{W}_{i,s}(\tau) = W_{i,s}(\tau) - \mathbb{E}(W_{i,s}(\tau)|S_i = s)$. There exist constants $0 < c < C < \infty$ such that

$$c < \inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon} \lambda_{\min}(\mathbb{E}\tilde{W}_{i,s}(\tau)\tilde{W}_{i,s}(\tau)^\top|S_i = s) \leq \sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon} \lambda_{\max}(\mathbb{E}\tilde{W}_{i,s}(\tau)\tilde{W}_{i,s}(\tau)^\top|S_i = s) \leq C$$

and $\mathbb{E}(\|\tilde{W}_{i,s}\|_2^d|S_i = s) \leq C$ for some $d > 2$, where for a generic symmetric matrix U, $\lambda_{\min}(U)$ and $\lambda_{\max}(U)$ denote the minimal and maximal eigenvalues of U, respectively.

The next theorem derives the closed-form expression for the optimal linear coefficient.

Theorem 5.2. *Suppose Assumptions 1, 2, 4, 6 and 7 hold, and $\Lambda_{\tau,s}(\cdot)$ is defined in (5.3). Further denote the asymptotic covariance matrix of $(\hat{q}^{par}(\tau_1), \ldots, \hat{q}^{par}(\tau_K))$ for any finite set of quantile indices $(\tau_1, \ldots, \tau_K)$ as $[\Sigma^{LP}(\tau_k, \tau_l)]_{k,l\in[K]}$. Then, $[\Sigma^{LP}(\tau_k, \tau_l)]_{k,l\in[K]}$ is minimized in the matrix sense at $\big(\theta_{1,s}(\tau_k), \theta_{0,s}(\tau_k)\big)_{k\in[K]}$ such that*

$$\frac{\theta_{1,s}(\tau_k)}{f_1(q_1(\tau_k))} + \frac{\pi(s)\theta_{0,s}(\tau_k)}{(1-\pi(s))f_0(q_0(\tau_k))} = \frac{\theta_{1,s}^{LP}(\tau_k)}{f_1(q_1(\tau_k))} + \frac{\pi(s)\theta_{0,s}^{LP}(\tau_k)}{(1-\pi(s))f_0(q_0(\tau_k))}, \quad k \in [K],$$

where for $\tau = \tau_1, \ldots, \tau_K$ and $a = 0, 1$,

$$\theta_{a,s}^{LP}(\tau) = \Big[\mathbb{E}(\tilde{W}_{i,s}(\tau)\tilde{W}_{i,s}(\tau)^\top|S_i = s)\Big]^{-1} \mathbb{E}\Big[\tilde{W}_{i,s}(\tau)1\{Y_i(a) \leq q_a(\tau)\}|S_i = s\Big].$$

Four remarks are in order. First, the optimal linear coefficients $\{\theta_{a,s}(\tau)\}_{a=0,1,s\in\mathcal{S}}$ are not uniquely defined. In order to achieve the minimal variance, we only need to consistently estimate one of the minimizers. We choose

$$(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) = (\theta_{1,s}^{LP}(\tau), \theta_{0,s}^{LP}(\tau)), \ \ s \in \mathcal{S},$$

as this choice avoids estimation of the densities $f_1(q_1(\tau))$ and $f_0(q_0(\tau))$. In Theorem 5.3 below, we propose estimators of $\theta_{1,s}^{LP}(\tau)$ and $\theta_{0,s}^{LP}(\tau)$ and show they are consistent uniformly over s and τ. Second, note that no adjustment is nested by our linear adjustment with zero coefficients. Due to the optimality result established in Theorem 5.2, our regression-adjusted QTE estimator with (consistent estimators of) $\{\theta_{a,s}^{LP}(\tau)\}_{a=0,1,s\in\mathcal{S}}$ is more efficient than that with no adjustments. Third, we also need to clarify that the optimality of $\{\theta_{a,s}^{LP}(\tau)\}_{a=0,1,s\in\mathcal{S}}$ is only within the class of linear regressions. It is possible that the QTE estimator with some nonlinear adjustments are more efficient than that with the optimal linear adjustments, especially because the linear probability model is likely misspecified. Fourth, the optimal linear coefficients $\{\theta_{a,s}^{LP}(\tau)\}_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon}$ minimize (over the class of linear models) not only the asymptotic variance of $\hat{q}^{par}(\tau)$ but also the covariance matrix of $(\hat{q}^{par}(\tau_1), \ldots, \hat{q}^{par}(\tau_K))$ for any finite-dimension quantile indices $(\tau_1, \ldots, \tau_K)$. This implies we can use the same (estimators of) optimal linear coefficients for hypothesis testing involving single, multiple, or even a continuum of quantile indices.

In the rest of this subsection, we focus on the estimation of $\{\theta_{a,s}^{LP}(\tau)\}_{a=0,1,s\in\mathcal{S}}$. Note that $\theta_{a,s}^{LP}(\tau)$ is the projection coefficient of $1\{Y_i \leq q_a(\tau)\}$ on $\tilde{W}_{i,s}(\tau)$ for the sub-population with $S_i = s$ and $A_i = a$. We estimate them by sample analog. Specifically, the parameter $q_a(\tau)$ is unknown and is replaced by some $\sqrt{n}$-consistent estimator denoted by $\hat{q}_a(\tau)$.

Assumption 8. Assume that $\sup_{\tau\in\Upsilon,a=0,1} |\hat{q}_a(\tau) - q_a(\tau)| = O_p(n^{-1/2})$.

In practice, we compute $\{\hat{q}_a(\tau)\}_{a=0,1}$ based on (3.1) and (3.2) by setting $\widehat{m}_a(\tau, S_i, X_i) \equiv 0$. Then, Assumption 8 holds automatically by Theorem 3.1 with $\widehat{m}_a(\tau, S_i, X_i) = \overline{m}_a(\tau, S_i, X_i) = 0$. Analysis throughout this section takes into account that the estimator $\hat{q}_a(\tau)$ is used in place of $q_a(\tau)$.

Next, we define the estimator of $\theta_{a,s}^{LP}(\tau)$. Recall $I_a(s)$ is defined in Assumption 3. Let

$$\overline{m}_a(\tau, s, X_i) = \tau - W_{i,s}^\top(\tau)\theta_{a,s}^{LP}(\tau), \tag{5.4}$$

$$\widehat{m}_a(\tau, s, X_i) = \tau - W_{i,s}^\top(\tau)\hat{\theta}_{a,s}^{LP}(\tau), \tag{5.5}$$

$$\dot{W}_{i,a,s}(\tau) = W_{i,s}(\tau) - \frac{1}{n_a(s)}\sum_{i \in I_a(s)} W_{i,s}(\tau), \tag{5.6}$$

and

$$\hat{\theta}_{a,s}^{LP}(\tau) = \left[\frac{1}{n_a(s)}\sum_{i \in I_a(s)} \dot{W}_{i,a,s}(\tau)\dot{W}_{i,a,s}^\top(\tau)\right]^{-1}\left[\frac{1}{n_a(s)}\sum_{i \in I_a(s)} \dot{W}_{i,a,s}(\tau)1\{Y_i \le \hat{q}_a(\tau)\}\right]. \tag{5.7}$$

Assumption 9. Suppose there exists a positive random variable L_i and a positive constant $C > 0$ such that for $a = 0, 1$,

$$\sup_{\tau_1,\tau_2 \in \Upsilon, a=0,1, s\in\mathcal{S}} \|W_{i,s}(\tau_1) - W_{i,s}(\tau_2)\|_2 \le L_i|\tau_1 - \tau_2|, \qquad \sup_{\tau\in\Upsilon, a=0,1, s\in\mathcal{S}} \|W_{i,s}(\tau)\|_2 \le L_i,$$

and $\mathbb{E}(L_i^d|S_i = s) \le C < \infty$ for some $d > 2$.

We note that Assumption 9 holds automatically if the regressor $W_{i,s}(\tau)$ does not depend on τ.

Theorem 5.3. *Suppose Assumptions 1, 2, 7–9 hold. Then Assumption 6 holds for* $(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau)) = (\theta_{a,s}^{LP}(\tau), \hat{\theta}_{a,s}^{LP}(\tau)),\ a = 0, 1, s \in \mathcal{S}, \tau \in \Upsilon$.

We refer to the QTE estimator adjusted by this linear probability model with optimal linear coefficients $\theta_{a,s}^{LP}(\tau)$ and estimators $\hat{\theta}_{a,s}^{LP}(\tau)$ as the LP estimator and denote it and its bootstrap counterpart as $\hat{q}^{LP}(\tau)$ and $\hat{q}^{LP,w}(\tau)$, respectively. Theorem 5.3 verifies Assumption 6 for the proposed estimator of the optimal linear coefficient. Then, by Theorem 5.1, Theorems 3.1 and 4.1 hold for $\hat{q}^{LP}(\tau)$ and $\hat{q}^{LP,w}(\tau)$, which implies all the estimation and inference methods established in the paper are valid for the LP estimator. Theorem 5.2 further shows $\hat{q}^{LP}(\tau)$ is the estimator with the optimal linear adjustment and weakly more efficient than the QTE estimator with no adjustments.

5.1.2. Logistic probability model

It is also common to consider the logistic regression as the adjustment and estimate the model by maximum likelihood (ML). The main goal of the working model is to approximate the true model as closely as possible. It is, therefore, useful to include additional technical regressors such as interactions in the logistic regression. The set of regressors used is defined as $H_i = H(X_i)$, which is allowed to contain the intercept. Let $\hat{\theta}_{a,s}^{ML}(\tau)$ and $\theta_{a,s}^{ML}(\tau)$ be the quasi-ML estimator and its corresponding pseudo true value, respectively, i.e.,

$$\hat{\theta}_{a,s}^{ML}(\tau) = \arg\max_{\theta_a} \frac{1}{n_a(s)}\sum_{i\in I_a(s)} \left[1\{Y_i \le \hat{q}_a(\tau)\}\log(\lambda(H_i^\top\theta_a)) + 1\{Y_i > \hat{q}_a(\tau)\}\log(1-\lambda(H_i^\top\theta_a))\right], \tag{5.8}$$

and

$$\theta_{a,s}^{ML}(\tau) = \arg\max_{\theta_a} \mathbb{E}\left[1\{Y_i(a) \le q_a(\tau)\}\log(\lambda(H_i^\top\theta_a)) + 1\{Y_i(a) > q_a(\tau)\}\log(1-\lambda(H_i^\top\theta_a))|S_i = s\right]. \tag{5.9}$$

We then define

$$\overline{m}_a(\tau, s, X_i) = \tau - \lambda(H_i^\top\theta_{a,s}^{ML}(\tau)) \quad \text{and} \quad \widehat{m}_a(\tau, s, X_i) = \tau - \lambda(H_i^\top\hat{\theta}_{a,s}^{ML}(\tau)). \tag{5.10}$$

In addition to the inclusion of technical regressors, we allow the pseudo true value $(\theta_{a,s}^{ML}(\tau))$ to vary across quantiles τ, giving another layer of flexibility to the model. Such a model is called the distribution regression and was first proposed by Chernozhukov et al. (2013). We emphasize here that, although we aim to make the regression model as flexible as possible, our theory and results do not require the model to be correctly specified.

Assumption 10. Suppose $\theta_{a,s}^{ML}(\tau)$ is the unique minimizer defined in (5.9) for $a = 0, 1$.

Theorem 5.4. *Suppose Assumptions 1, 2, 8, 10 hold and there exist constants c, C such that*

$$0 < c \le \lambda_{\min}(\mathbb{E}H_iH_i^\top) \le \lambda_{\max}(\mathbb{E}H_iH_i^\top) \le C < \infty,$$

then Assumption 6(iii) holds for $(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau)) = (\theta_{a,s}^{ML}(\tau), \hat{\theta}_{a,s}^{ML}(\tau)),\ a = 0, 1, s \in \mathcal{S}, \tau \in \Upsilon$.

Four remarks are in order. First, we refer to the QTE estimator adjusted by the logistic model with QMLE as the ML estimator and denote it and its bootstrap counterpart as $\hat{q}^{ML}(\tau)$ and $\hat{q}^{ML,w}(\tau)$, respectively. Assumption 6(i) holds automatically for the logistic regression. If we further impose Assumption 6(ii), then Theorem 5.4 implies that all the estimation and bootstrap inference methods established in the paper are valid for the ML estimator. Second, we take into account that $\hat{\theta}_{a,s}^{ML}(\tau)$ is computed when the true $q_a(\tau)$ is replaced by its estimator $\hat{q}_a(\tau)$ and derive the results in

Theorem 5.4 under Assumption 8. Third, the ML estimator is not guaranteed to be optimal or more efficient than QTE estimator with no adjustments. On the other hand, as we can include additional technical terms in the regression and allow the regression coefficients to vary across τ, the logistic model can be close to the true model $m_a(\tau, s, X_i)$, which achieves the global minimum asymptotic variance based on Theorem 3.1. Fourth, in Section 5.2, we further justify the use of the ML estimator with a flexible logistic model by letting the number of technical terms (or equivalently, the dimension of H_i) diverge to infinity, showing by this means that the ML estimator can indeed consistently estimate the true model and thereby achieve the global minimum covariance matrix of the adjusted QTE estimator.

5.1.3. Further improved logistic model

Although in simulations, we cannot find a DGP in which the QTE estimator with logistic adjustment is less efficient than that with no adjustments, theoretically such a scenario still exists. In this section, we follow the idea of Cohen and Fogarty (2020) and construct an estimator which is weakly more efficient than both the ML estimator and the estimator with no adjustments. We denote $W_{i,s}(\tau) = (\lambda(H_i^\top \theta_{1,s}^{ML}(\tau)), \lambda(H_i^\top \theta_{0,s}^{ML}(\tau)))^\top$ and treat it as the regressor in a linear adjustment, i.e., define $\overline{m}_a(\tau, s, X_i) = \tau - W_{i,s}^\top(\tau)t_{a,s}(\tau)$. Then, the logistic adjustment in Section 5.1.2 and no adjustments correspond to $t_{a,s}(\tau) = a(1, 0)^\top + (1-a)(0, 1)^\top$ and $t_{a,s}(\tau) = (0, 0)^\top$ for $a = 0, 1$, respectively. However, following Theorem 5.2, the optimal linear coefficient with regressor $W_{i,s}(\tau)$ is

$$\theta_{a,s}^{LPML}(\tau) = \left[\mathbb{E}(\tilde{W}_{i,s}(\tau)\tilde{W}_{i,s}^\top(\tau)|S_i = s)\right]^{-1} \mathbb{E}\left[\tilde{W}_{i,s}(\tau)1\{Y_i(a) \leq q_a(\tau)\}|S_i = s\right], \tag{5.11}$$

where $\tilde{W}_{i,s}(\tau) = W_{i,s}(\tau) - \mathbb{E}(W_{i,s}(\tau)|S_i = s)$. Using the adjustment term $\overline{m}_a(\tau, s, X_i) = \tau - W_{i,s}^\top(\tau)t_{a,s}(\tau)$ with $t_{a,s}(\tau) = \theta_{a,s}^{LPML}(\tau)$ is asymptotically weakly more efficient than any other choices of $t_{a,s}(\tau)$. In practice, we do not observe $W_{i,s}(\tau)$, but can replace it by its feasible version $\hat{W}_{i,s}(\tau) = (\lambda(H_i^\top \hat{\theta}_{1,s}^{ML}(\tau)), \lambda(H_i^\top \hat{\theta}_{0,s}^{ML}(\tau)))^\top$. We then define

$$\overline{m}_a(\tau, s, X_i) = \tau - W_{i,s}^\top(\tau)\theta_{a,s}^{LPML}(\tau), \tag{5.12}$$

$$\widehat{m}_a(\tau, s, X_i) = \tau - \hat{W}_{i,s}^\top(\tau)\hat{\theta}_{a,s}^{LPML}(\tau), \tag{5.13}$$

$$\breve{W}_{i,a,s}(\tau) = \hat{W}_{i,s}(\tau) - \frac{1}{n_a(s)}\sum_{i \in I_a(s)} \hat{W}_{i,s}(\tau), \tag{5.14}$$

and

$$\hat{\theta}_{a,s}^{LPML}(\tau) = \left[\frac{1}{n_a(s)}\sum_{i \in I_a(s)} \breve{W}_{i,a,s}(\tau)\breve{W}_{i,a,s}^\top(\tau)\right]^{-1}\left[\frac{1}{n_a(s)}\sum_{i \in I_a(s)} \breve{W}_{i,a,s}(\tau)1\{Y_i \leq \hat{q}_a(\tau)\}\right]. \tag{5.15}$$

Assumption 11.

(i) There exist constants c, C such that

$$0 < c < \inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon} \lambda_{\min}(\mathbb{E}(\tilde{W}_{i,s}(\tau)\tilde{W}_{i,s}^\top(\tau)|S_i = s)) \leq \sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon} \lambda_{\max}(\mathbb{E}(\tilde{W}_{i,s}(\tau)\tilde{W}_{i,s}^\top(\tau)|S_i = s)) \leq C < \infty.$$

(ii) Suppose

$$\sup_{\tau_1,\tau_2\in\Upsilon,a=0,1,s\in\mathcal{S}} \|\theta_{a,s}^{ML}(\tau_1) - \theta_{a,s}^{ML}(\tau_2)\|_2 \leq C|\tau_1 - \tau_2|$$
$$\sup_{\tau_1,\tau_2\in\Upsilon,a=0,1,s\in\mathcal{S}} \|\theta_{a,s}^{LPML}(\tau_1) - \theta_{a,s}^{LPML}(\tau_2)\|_2 \leq C|\tau_1 - \tau_2|.$$

Theorem 5.5. *Denote $\hat{q}^{LPML}(\tau)$ and $\hat{q}^{LPML,w}(\tau)$ as the τth QTE estimator and its multiplier bootstrap counterpart defined in Sections 3 and 4, respectively, with $\overline{m}_a(\tau, s, X_i)$ and $\widehat{m}_a(\tau, s, X_i)$ defined in (5.12) and (5.13), respectively. Suppose Assumptions 1, 2, 8, 10 and 11 hold, and there exist constants c, C such that*

$$0 < c \leq \lambda_{\min}(\mathbb{E}H_iH_i^\top) \leq \lambda_{\max}(\mathbb{E}H_iH_i^\top) \leq C < \infty.$$

Then, Assumptions 3 and 5 hold, which further implies Theorems 3.1 and 4.1 hold for $\hat{q}^{LPML}(\tau)$ and $\hat{q}^{LPML,w}(\tau)$, respectively. Further denote the asymptotic covariance matrices of $(\hat{q}^J(\tau_1), \ldots, \hat{q}^J(\tau_K))$ for any finite set of quantile indices $(\tau_1, \ldots, \tau_K)$ as $[\Sigma^J(\tau_k, \tau_l)]_{k,l\in[K]}$ for $J \in \{LPML, ML, NA\}$, where $\hat{q}^{NA}(\tau)$ is the τth QTE estimator without adjustments. Then we have

$$[\Sigma^{LPML}(\tau_k, \tau_l)]_{k,l\in[K]} \leq [\Sigma^{ML}(\tau_k, \tau_l)]_{k,l\in[K]} \quad \text{and} \quad [\Sigma^{LPML}(\tau_k, \tau_l)]_{k,l\in[K]} \leq [\Sigma^{NA}(\tau_k, \tau_l)]_{k,l\in[K]}$$

in the matrix sense.

In practice when n is small $\breve{W}_{i,a,s}(\tau)$ may be nearly multicollinear within some stratum, which can lead to size distortion in inference concerning QTE. We therefore suggest first normalizing $\breve{W}_{i,a,s}(\tau)$ by its standard deviation (denoting the normalized $\breve{W}_{i,a,s}(\tau)$ as $\ddot{W}_{i,a,s}(\tau)$) and then running a ridge regression

$$\tilde{\theta}^{LPML}_{a,s}(\tau) = \left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)}\ddot{W}_{i,a,s}(\tau)\ddot{W}^\top_{i,a,s}(\tau) + \delta_n I_2\right]^{-1}\left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)}\ddot{W}_{i,a,s}(\tau)1\{Y_i \le \hat{q}_a(\tau)\}\right],$$

where I_2 is the two-dimensional identity matrix and $\delta_n = 1/n$. Then, the final regression adjustment is

$$\widehat{m}_a(\tau, s, X_i) = \tau - \ddot{W}^\top_{i,a,s}(\tau)\tilde{\theta}^{LPML}_{a,s}(\tau).$$

Given Assumption 11, such a ridge penalty is asymptotically negligible and all the results in Theorem 5.5 still hold.[7]

5.2. Nonparametric method

This section considers nonparametric estimation of $m_a(\tau, s, X_i)$ when the dimension of X_i is fixed as d_x. For ease of notation, we assume all coordinates of X_i are continuously distributed. If in an application some elements of X are discrete, the dimension d_x is interpreted as the dimension of the continuous covariates. All results in this section can then be extended in a conceptually straightforward manner by using the continuous covariates only within samples that are homogeneous in discrete covariates.

As $m_a(\tau, s, X_i)$ is nonparametrically estimated, we have $\overline{m}_a(\tau, s, X_i) = m_a(\tau, s, X_i) = \tau - \mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i)$. We estimate $\mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i)$ by the sieve method of fitting a logistic model, as studied by Hirano et al. (2003). Specifically, recall $\lambda(\cdot)$ is the logistic CDF and denote the number of sieve bases by h_n, which depends on the sample size n and can grow to infinity as $n \to \infty$. Let $H_{h_n}(x) = (b_{1n}(x), \ldots, b_{h_n n}(x))^\top$ where $\{b_{hn}(x)\}_{h\in[h_n]}$ is an h_n dimensional basis of a linear sieve space. More details on the sieve space are given in Section B of the Online Supplement. Denote

$$\widehat{m}_a(\tau, s, X_i) = \tau - \lambda(H^\top_{h_n}(X_i)\hat{\theta}^{NP}_{a,s}(\tau)) \quad \text{and} \tag{5.16}$$

$$\hat{\theta}^{NP}_{a,s}(\tau) = \arg\max_{\theta_a}\frac{1}{n_a(s)}\sum_{i\in I_a(s)}\Big[1\{Y_i \le \hat{q}_a(\tau)\}\log(\lambda(H^\top_{h_n}(X_i)\theta_a)) + 1\{Y_i > \hat{q}_a(\tau)\}\log(1-\lambda(H^\top_{h_n}(X_i)\theta_a))\Big]. \tag{5.17}$$

We refer to the QTE estimator with the nonparametric adjustment as the NP estimator. Note that we use the estimator $\hat{q}_a(\tau)$ of $q_a(\tau)$ in (5.17), where $\hat{q}_a(\tau)$ satisfies Assumption 8. All the analysis in this section takes account of the fact that $\hat{q}_a(\tau)$ instead of $q_a(\tau)$ is used.

Assumption 12.

(i) There exist constants $0 < \kappa_1 < \kappa_2 < \infty$ such that with probability approaching one,

$$\kappa_1 \le \lambda_{\min}\left(\frac{1}{n_a(s)}\sum_{i\in I_a(s)}H_{h_n}(X_i)H^\top_{h_n}(X_i)\right) \le \lambda_{\max}\left(\frac{1}{n_a(s)}\sum_{i\in I_a(s)}H_{h_n}(X_i)H^\top_{h_n}(X_i)\right) \le \kappa_2,$$

and

$$\kappa_1 \le \lambda_{\min}\left(\mathbb{E}(H_{h_n}(X_i)H^\top_{h_n}(X_i)|S_i = s)\right) \le \lambda_{\max}\left(\mathbb{E}(H_{h_n}(X_i)H^\top_{h_n}(X_i)|S_i = s)\right) \le \kappa_2.$$

(ii) For $a = 0, 1$, there exists an $h_n \times 1$ vector $\theta^{NP}_{a,s}(\tau)$ such that for $R_a(\tau, s, x) = \mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i = x) - \lambda(H^\top_{h_n}(x)\theta^{NP}_{a,s}(\tau))$, we have $\sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\mathrm{Supp}(X)}|R_a(\tau, s, x)| = o(1)$,

$$\sup_{a=0,1,\tau\in\Upsilon,s\in\mathcal{S}}\frac{1}{n_a(s)}\sum_{i\in I_a(s)}R^2_a(\tau, s, X_i) = O_p\left(\frac{h_n\log(n)}{n}\right),$$

and

$$\sup_{a=0,1,\tau\in\Upsilon,s\in\mathcal{S}}\mathbb{E}(R^2_a(\tau, s, X_i)|S_i = s) = O\left(\frac{h_n\log(n)}{n}\right).$$

(iii) For $a = 0, 1$, there exists a constant $c \in (0, 0.5)$ such that

$$c \le \inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\mathrm{Supp}(X)}\mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i = x) \le \sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\mathrm{Supp}(X)}\mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i = x) \le 1 - c.$$

[7] In unreported simulations, we find that when $n \ge 800$, the ridge regularization is unnecessary and the original adjustment (i.e., (5.13)) has no size distortion, implying that near-multicollinearity is indeed just a finite-sample issue.

(iv) Suppose $\mathbb{E}(H^2_{h_n,h}(X_i)|S_i = s) \leq C < \infty$ for some constant $C > 0$, $\sup_{x\in \mathrm{Supp}(X)} \|H_{h_n}(x)\|_2 \leq \zeta(h_n)$, $\zeta^2(h_n)h_n \log(n) = o(n)$, and $h_n^2 \log^2(n) = o(n)$, where $H_{h_n,h}(X_i)$ denotes the hth coordinate of $H_{h_n}(X_i)$.

Four remarks are in order. First, Assumption 12(i) is standard in the sieve literature. Second, Assumption 12(ii) means the approximation error of the sieve logistic model vanishes asymptotically, which holds given sufficient smoothness of $\mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x)$ in x. Third, Assumption 12(iii) usually holds when $\mathrm{Supp}(X)$ is compact. This condition is also assumed by Hirano et al. (2003). Fourth, the quantity $\zeta(h_n)$ in Assumption 12(iv) depends on the choice of basis functions. For example, $\zeta(h_n) = O(h_n^{1/2})$ for splines and $\zeta(h_n) = O(h_n)$ for power series. Taking splines as an example, Assumption 12(iv) requires $h_n = o(n^{1/2})$.

Theorem 5.6. *Denote $\hat{q}^{NP}(\tau)$ and $\hat{q}^{NP,w}(\tau)$ as the τth QTE estimator and its multiplier bootstrap counterpart defined in Sections 3 and 4, respectively, with $\overline{m}_a(\tau, S_i, X_i) = m_a(\tau, S_i, X_i)$ and $\widehat{m}_a(\tau, S_i, X_i)$ defined in (5.16). Further suppose Assumption 1, 2, 4, 8, and 12 hold. Then, Assumptions 3 and 5 hold, which further implies that Theorems 3.1 and 4.1 hold for $\hat{q}^{NP}(\tau)$ and $\hat{q}^{NP,w}(\tau)$, respectively. In addition, for any finite-dimensional quantile indices $(\tau_1, \ldots, \tau_K)$, the covariance matrix of $(\hat{q}^{NP}(\tau_1), \ldots, \hat{q}^{NP}(\tau_K))$ achieves the minimum (in the matrix sense) as characterized in Theorem 3.1.*

Three remarks are in order. First, as the nonparametric regression consistently estimates the true specifications $\{m_a(\cdot)\}_{a=0,1}$, the QTE estimator adjusted by the nonparametric regression achieves the global minimum asymptotic variance, and thus is weakly more efficient than QTE estimation with linear and logistic adjustments studied in the previous section. Second, the practical implementation of NP and ML methods are the same, given that they share the same set of covariates (basis functions). Therefore, even if we include a small number of basis functions so that h_n is better treated as fixed, the proposed estimation and inference methods for the regression-adjusted QTE estimator are still valid, although they may not be optimal. Third, in Section A in the Online Supplement, we consider computing $\widehat{m}_a(\tau, s, x)$ via an ℓ_1 penalized logistic regression when the dimension of the regressors can be comparable or even higher than the sample size. We then provide primitive conditions under which we verify Assumptions 3 and 5.

6. Simulations

6.1. Data generating processes

Two DGPs are used to assess the finite sample performance of the estimation and inference methods introduced in the paper. We consider the outcome equation

$$Y_i = \alpha(X_i) + \gamma Z_i + \mu(X_i)A_i + \eta_i, \tag{6.1}$$

where $\gamma = 4$ for all cases while $\alpha(X_i)$, $\mu(X_i)$, and η_i are separately specified as follows.

(i) Let Z be standardized Beta(2, 2) distributed, $S_i = \sum_{j=1}^4 1\{Z_i \leq g_j\}$, and $(g_1, \ldots, g_4) = (-0.25\sqrt{20}, 0, 0.25\sqrt{20}, 0.5\sqrt{20})$. X_i contains two covariates $(X_{1i}, X_{2i})^\top$, where X_{1i} follows a uniform distribution on $[-2, 2]$, X_{2i} follows a standard normal distribution, and X_{1i} and X_{2i} are independent. Further define $\alpha(X_i) = 1 + X_{2i}$, $\mu(X_i) = 1 + X_i^\top\beta$, $\beta = (3, 3)^\top$, and $\eta_i = (0.25 + X_{1i}^2)A_i\varepsilon_{1i} + (1 - A_i)\varepsilon_{2i}$, where $(\varepsilon_{1i}, \varepsilon_{2i})$ are jointly standard normal.

(ii) Let Z be uniformly distributed on $[-2, 2]$, $S_i = \sum_{j=1}^4 1\{Z_i \leq g_j\}$, and $(g_1, \ldots, g_4) = (-1, 0, 1, 2)$. Let $X_i = (X_{1i}, X_{2i})^\top$ be the same as defined in DGP (i). Further define $\alpha(X_i) = 1 + X_{1i} + X_{2i}$, $\mu(X_i) = 1 + X_{1i} + X_{2i} + \frac{1}{4}(X_i^\top\beta)^2$ with $\beta = (2, 2)^\top$, and $\eta_i = 2(1 + Z_i^2)A_i\varepsilon_{1i} + (1 + Z_i^2)(1 - A_i)\varepsilon_{2i}$, where $(\varepsilon_{1i}, \varepsilon_{2i})$ are mutually independently $T(5)/\sqrt{5}$ distributed.

For each DGP, we consider the following four randomization schemes as in Zhang and Zheng (2020) with $\pi(s) = 0.5$ for $s \in \mathcal{S}$:

(i) SRS: Treatment assignment is generated as in Example 1.
(ii) WEI: Treatment assignment is generated as in Example 2 with $\phi(x) = (1 - x)/2$.
(iii) BCD: Treatment assignment is generated as in Example 3 with $\lambda = 0.75$.
(iv) SBR: Treatment assignment is generated as in Example 4.

We assess the empirical size and power of the tests for $n = 200$ and $n = 400$. We compute the true QTEs and their differences by simulations with 10,000 sample size and 1,000 replications. To compute power, we perturb the true values by $\Delta = 1.5$. We examine three null hypotheses:

(i) Pointwise test

$$H_0 : q(\tau) = \text{truth} \quad \text{v.s.} \quad H_1 : q(\tau) = \text{truth} + \Delta, \quad \tau = 0.25, 0.5, 0.75;$$

(ii) Test for the difference

$$H_0 : q(0.75) - q(0.25) = \text{truth} \quad \text{v.s.} \quad H_1 : q(0.75) - q(0.25) = \text{truth} + \Delta;$$

(iii) Uniform test

$$H_0 : q(\tau) = \text{truth}(\tau) \quad v.s. \quad H_1 : q(\tau) = \text{truth}(\tau) + \Delta, \quad \tau \in [0.25, 0.75].$$

For the pointwise test, we report the results for the median ($\tau = 0.5$) in the main text and give the cases $\tau = 0.25$ and $\tau = 0.75$ in the Online Supplement.

6.2. Estimation methods

We consider the following estimation methods of the auxiliary regression.

(i) NA: the estimator with no adjustments, i.e., setting $\widehat{m}_a(\cdot) = \overline{m}_a(\cdot) = 0$.
(ii) LP: the linear probability model with regressors X_i and the pseudo true value estimated by $\hat{\theta}^{LP}_{a,s}(\tau)$ defined in (5.7).
(iii) ML: the logistic model with regressor $H_i = (1, X_{1i}, X_{2i})^\top$ and the pseudo true value estimated by $\hat{\theta}^{ML}_{a,s}(\tau)$ defined in (5.8).
(iv) LPML: the logistic model with regressor $H_i = (1, X_{1i}, X_{2i})^\top$ and the pseudo true value estimated by $\hat{\theta}^{LPML}_{a,s}(\tau)$ defined in (5.15).
(v) MLX: the logistic model with regressor $H_i = (1, X_{1i}, X_{2i}, X_{1i}X_{2i})^\top$ and the pseudo true value estimated by $\hat{\theta}^{ML}_{a,s}(\tau)$ defined in (5.8).
(vi) LPMLX: the logistic model with regressor $H_i = (1, X_{1i}, X_{2i}, X_{1i}X_{2i})^\top$ and the pseudo true value estimated by $\hat{\theta}^{LPML}_{a,s}(\tau)$ defined in (5.15).
(vii) NP: the logistic model with regressor $H_{h_n}(X_i) = (1, X_{1i}, X_{2i}, X_{1i}X_{2i}, X_{1i}1\{X_{1i} > t_1\}X_{2i}1\{X_{2i} > t_2\})^\top$ where t_1 and t_2 are the sample medians of $\{X_{1i}\}_{i\in[n]}$ and $\{X_{2i}\}_{i\in[n]}$, respectively. The pseudo true value is estimated by $\hat{\theta}^{NP}_{a,s}(\tau)$ defined in (5.17).

6.3. Simulation results

Table 1 presents the empirical size and power for the pointwise test with $\tau = 0.5$ under DGPs (i) and (ii). We make six observations. First, none of the auxiliary regressions is correctly specified, but test sizes are all close to the nominal level 5%, confirming that estimation and inference are robust to misspecification. Second, the inclusion of auxiliary regressions improves the efficiency of the QTE estimator as the powers for method "NA" are the lowest among all the methods for both DGPs and all randomization schemes. This finding is consistent with theory because methods "LP", "LPML", "LPMLX", "NP" are guaranteed to be weakly more efficient than "NA". Third, the powers of methods "LPML" and "LPMLX" are higher than those of methods "ML" and "MLX", respectively. This is consistent with our theory that methods "LPML" and "LPMLX" further improve "ML" and "MLX", respectively. In addition, methods "MLX" and "LPMLX" fit a flexible distribution regression that can approximate the true DGP well. Therefore, the powers of "MLX" and "LPMLX" are respectively much larger than those of "ML" and "LPML". For the same reason we observe that the power of "LPMLX" is close to "NP".[8] Fourth, the powers of method "NP" are the best because it estimates the true specification and achieves the minimum asymptotic variance of $\hat{q}^{adj}(\tau)$ as shown in Theorem 5.1. Fifth, when the sample size is 200, the method "NP" slightly over-rejects but size becomes closer to nominal when the sample size increases to 400. Sixth, the improvement of power of "LPMLX" estimator upon "NA" (i.e., with no adjustments) is due to about 12%–15% reduction of the standard error of the QTE estimator on average.[9]

Tables 2 and 3 present sizes and powers of inference on $q(0.75) - q(0.25)$ and on $q(\tau)$ uniformly over $\tau \in [0.25, 0.75]$, respectively, for DGPs (i) and (ii) and four randomization schemes. All the observations made above apply to these results. The improvement in power of the "LPMLX" estimator upon "NA" (i.e., with no adjustments) is due to a 9% reduction of the standard error of the difference of the QTE estimators on average. In Section C in the Online Supplement, we provide additional simulation results such as the empirical sizes and powers for the pointwise test with $\tau = 0.25$ and 0.75, the bootstrap inference with the true target fraction, and the adjusted QTE estimator when the DGP contains high-dimensional covariates and the adjustments are computed via logistic Lasso. We also report the biases and standard errors of the adjusted QTE estimators.

6.4. Practical recommendations

When X is finite-dimensional, we suggest using the LPMLX adjustment in which the logistic model includes interaction terms and the regression coefficients are allowed to depend on (τ, a, s). When X is high-dimensional, we suggest using the logistic Lasso to estimate the regression adjustment.[10]

[8] The results in Section C of the Online Supplement show that "LPMLX" has much smaller bias than "NP" and its variance is similar to "NP", which make "LPMLX" preferable in practice.

[9] The bias and standard errors are reported in Section C in the Online Supplement.

[10] The relevant theory and simulation results on high-dimensional covariates are provided in Section A of the Online Supplement.

Table 1
Pointwise test ($\tau = 0.5$).

Methods	Size								Power							
	$N = 200$				$N = 400$				$N = 200$				$N = 400$			
	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
Panel A: DGP (i)																
NA	0.055	0.054	0.050	0.054	0.051	0.054	0.051	0.051	0.404	0.406	0.403	0.406	0.665	0.676	0.681	0.681
LP	0.052	0.050	0.049	0.052	0.048	0.053	0.051	0.052	0.491	0.497	0.502	0.492	0.779	0.788	0.790	0.791
ML	0.053	0.050	0.049	0.055	0.051	0.050	0.052	0.052	0.472	0.478	0.483	0.473	0.759	0.768	0.775	0.773
LPML	0.054	0.052	0.052	0.057	0.052	0.054	0.051	0.053	0.506	0.509	0.523	0.513	0.802	0.812	0.814	0.809
MLX	0.056	0.059	0.055	0.057	0.055	0.054	0.055	0.058	0.475	0.479	0.486	0.482	0.752	0.759	0.760	0.760
LPMLX	0.060	0.058	0.059	0.058	0.054	0.055	0.054	0.054	0.506	0.513	0.521	0.512	0.802	0.810	0.813	0.811
NP	0.063	0.059	0.062	0.064	0.055	0.054	0.054	0.056	0.523	0.523	0.531	0.526	0.804	0.811	0.814	0.809
Panel B: DGP (ii)																
NA	0.046	0.051	0.045	0.047	0.047	0.045	0.048	0.047	0.479	0.489	0.500	0.490	0.773	0.775	0.774	0.782
LP	0.049	0.051	0.050	0.050	0.045	0.048	0.050	0.045	0.572	0.581	0.589	0.579	0.851	0.856	0.857	0.854
ML	0.051	0.058	0.050	0.054	0.049	0.046	0.050	0.048	0.524	0.534	0.541	0.539	0.812	0.810	0.807	0.807
LPML	0.051	0.058	0.054	0.053	0.050	0.049	0.053	0.047	0.574	0.581	0.588	0.580	0.862	0.863	0.863	0.863
MLX	0.058	0.059	0.056	0.059	0.051	0.049	0.051	0.050	0.566	0.574	0.583	0.573	0.826	0.824	0.827	0.827
LPMLX	0.057	0.062	0.057	0.060	0.052	0.050	0.053	0.052	0.615	0.620	0.630	0.627	0.878	0.878	0.880	0.879
NP	0.063	0.066	0.062	0.062	0.056	0.055	0.056	0.051	0.622	0.625	0.632	0.628	0.883	0.880	0.882	0.879

Table 2
Test for differences ($\tau_1 = 0.25$, $\tau_2 = 0.75$).

Methods	Size								Power							
	$N = 200$				$N = 400$				$N = 200$				$N = 400$			
	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
Panel A: DGP (i)																
NA	0.043	0.045	0.040	0.041	0.044	0.043	0.041	0.043	0.214	0.216	0.209	0.203	0.387	0.389	0.383	0.365
LP	0.045	0.048	0.043	0.045	0.045	0.047	0.043	0.045	0.246	0.242	0.234	0.248	0.424	0.422	0.422	0.421
ML	0.045	0.045	0.043	0.042	0.046	0.047	0.040	0.048	0.234	0.233	0.231	0.239	0.415	0.422	0.417	0.426
LPML	0.044	0.049	0.045	0.045	0.049	0.049	0.044	0.047	0.250	0.250	0.248	0.259	0.451	0.453	0.450	0.459
MLX	0.046	0.052	0.046	0.047	0.047	0.047	0.044	0.049	0.232	0.234	0.229	0.241	0.415	0.415	0.404	0.416
LPMLX	0.049	0.055	0.047	0.047	0.049	0.050	0.047	0.047	0.247	0.249	0.249	0.258	0.445	0.453	0.445	0.453
NP	0.050	0.054	0.050	0.051	0.052	0.052	0.047	0.048	0.246	0.248	0.245	0.257	0.444	0.444	0.442	0.450
Panel B: DGP (ii)																
NA	0.039	0.044	0.040	0.038	0.044	0.041	0.039	0.047	0.211	0.225	0.217	0.194	0.399	0.396	0.392	0.383
LP	0.043	0.048	0.045	0.040	0.045	0.044	0.042	0.047	0.244	0.255	0.251	0.245	0.447	0.440	0.441	0.455
ML	0.049	0.046	0.046	0.043	0.044	0.045	0.042	0.048	0.217	0.228	0.213	0.212	0.379	0.386	0.386	0.396
LPML	0.047	0.051	0.048	0.043	0.047	0.045	0.047	0.048	0.253	0.258	0.253	0.252	0.456	0.451	0.454	0.468
MLX	0.047	0.051	0.047	0.047	0.046	0.046	0.045	0.049	0.226	0.240	0.228	0.223	0.394	0.392	0.391	0.399
LPMLX	0.053	0.056	0.051	0.048	0.051	0.049	0.045	0.050	0.261	0.272	0.265	0.263	0.467	0.460	0.460	0.477
NP	0.056	0.058	0.053	0.052	0.051	0.052	0.045	0.050	0.266	0.275	0.266	0.270	0.469	0.459	0.461	0.479

7. Empirical application

Undersaving has been found to have important individual and social welfare consequences (Karlan et al., 2014). Does expanding access to bank accounts for the poor lead to an overall increase in savings? To answer the question, Dupas et al. (2018) conducted a covariate-adaptive randomized experiment in Uganda, Malawi, and Chile to study the impact of a bank account subsidy on savings. In their paper, the authors examined the ATEs as well as the QTEs of the subsidy. This section reports an application of our methods to the same dataset to examine the QTEs of the subsidy on household total savings in Uganda.

The sample consists of 2160 households in Uganda.[11] Within each of 41 strata by gender, occupation, and bank branch, 50 percent of households in the sample were randomly assigned to receive the bank account subsidy and the rest of the sample were in the control group. This is a block stratified randomization design with 41 strata, which satisfies Assumption 1 in Section 2. The target fraction of the treated units is 1/2. It is trivial to see that statements (i), (ii), and (iii) in Assumption 1 are satisfied. Because $\max_{s\in\mathcal{S}} |\frac{D_n(s)}{n(s)}| \approx 0.056$, it is reasonable to claim that Assumption 1(iv) is also satisfied in our analysis.

[11] We filter out observations with missing values. Our final sample contains 1952 households.

Table 3
Uniform test ($\tau \in [0.25, 0.75]$).

Methods	Size								Power							
	$N = 200$				$N = 400$				$N = 200$				$N = 400$			
	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
Panel A: DGP (i)																
NA	0.048	0.044	0.044	0.045	0.047	0.049	0.045	0.048	0.450	0.451	0.455	0.454	0.765	0.770	0.769	0.770
LP	0.045	0.044	0.043	0.045	0.047	0.051	0.047	0.046	0.589	0.588	0.589	0.581	0.902	0.901	0.904	0.900
ML	0.047	0.044	0.043	0.045	0.044	0.051	0.045	0.047	0.570	0.577	0.582	0.568	0.887	0.889	0.893	0.890
LPML	0.046	0.046	0.045	0.047	0.046	0.050	0.046	0.051	0.603	0.605	0.616	0.607	0.916	0.917	0.915	0.915
MLX	0.052	0.049	0.048	0.048	0.046	0.053	0.050	0.050	0.582	0.582	0.595	0.576	0.889	0.893	0.891	0.889
LPMLX	0.053	0.047	0.049	0.052	0.047	0.053	0.050	0.050	0.612	0.614	0.619	0.610	0.915	0.919	0.919	0.913
NP	0.056	0.055	0.054	0.055	0.050	0.057	0.052	0.054	0.633	0.627	0.633	0.629	0.916	0.919	0.918	0.915
Panel B: DGP (ii)																
NA	0.038	0.039	0.039	0.038	0.045	0.039	0.040	0.045	0.572	0.571	0.579	0.574	0.878	0.882	0.879	0.879
LP	0.041	0.044	0.045	0.041	0.044	0.043	0.039	0.042	0.704	0.708	0.710	0.700	0.953	0.955	0.956	0.955
ML	0.044	0.043	0.048	0.041	0.047	0.045	0.043	0.044	0.661	0.660	0.664	0.655	0.931	0.931	0.933	0.935
LPML	0.047	0.046	0.048	0.044	0.047	0.046	0.041	0.046	0.723	0.714	0.720	0.714	0.964	0.963	0.965	0.964
MLX	0.052	0.050	0.052	0.049	0.048	0.046	0.045	0.045	0.703	0.710	0.708	0.704	0.946	0.949	0.946	0.951
LPMLX	0.056	0.054	0.054	0.051	0.052	0.048	0.046	0.048	0.761	0.761	0.766	0.754	0.972	0.972	0.972	0.974
NP	0.060	0.060	0.062	0.058	0.055	0.052	0.047	0.051	0.770	0.771	0.773	0.765	0.973	0.974	0.972	0.974

Table 4
QTEs on total savings (one auxiliary regressor).

	NA	LP	ML	LPML
25%	1.105	1.105	1.105	1.105
	(0.564)	(0.564)	(0.470)	(0.470)
50%	3.682	3.682	3.682	3.682
	(1.010)	(1.080)	(1.146)	(1.033)
75%	7.363	9.204	9.204	9.204
	(3.757)	(4.227)	(3.616)	(3.757)

Notes: The table presents the QTE estimates of the effect of the bank account subsidy on household total savings at quantiles 25%, 50%, and 75% when only one auxiliary regressor (baseline value of total savings) is used in the regression adjustment models. Standard errors are in parentheses.

After the randomization and the intervention, the authors conducted 3 rounds of follow-up surveys in Uganda (See Dupas et al. (2018) for a detailed description). In this section, we focus on the first round follow up survey to examine the impact of the bank account subsidy on total savings.

Tables 4 and 5 present the QTE estimates and their standard errors (in parentheses) estimated by different methods at quantile indices 0.25, 0.5, and 0.75. The description of these estimators is similar to that in Section 6.[12] In the analysis, we focus on two sets of additional baseline variables: baseline value of total savings only (one auxiliary regressor) and baseline value of total savings, household size, age, and married female dummy (four auxiliary regressors). The first set of regressors follows Dupas et al. (2018). The second one is used to illustrate all the methods discussed in the paper. Tables 4 and 5 report the results with one and four auxiliary regressors, respectively.

The results in Tables 4–5 prompt two observations. First, consistent with the theoretical and simulation results, the standard errors for the regression-adjusted QTEs are mostly lower than those for the QTE estimate without adjustment.

12 Specifically, we have:
(i) NA: the estimator with no adjustments.
(ii) LP: the linear probability model. When there is only one auxiliary regressor, $H_i = (1, X_{1i})^\top$, and when there are four auxiliary regressors, $H_i = (1, X_{1i}, X_{2i}, X_{3i}, X_{4i})^\top$, where $X_{1i}, X_{2i}, X_{3i}, X_{4i}$, represent four covariates used in the regression adjustment.
(iii) ML: the logistic probability model with regressor H_i, where H_i is the same as that in the LP model.
(iv) LPML: the further improved logistic probability model with regressor H_i, where H_i is the same as that in the LP model.
(v) MLX: the logistic probability model with interaction terms. MLX is only be applied to the case when there are four auxiliary regressors with $H_i = (1, X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{1i}X_{2i}, X_{2i}X_{3i})^\top$.
(vi) LPMLX: the further improved logistic probability model with interaction terms. LPMLX is only be applied to the case when there are four auxiliary regressors with the same H_i as that used in the MLX model.
(vii) NP: the nonparametric logistic probability model with regressor H_{h_n}. NP is only applied to the case when there are four auxiliary regressors with $H_{h_n} = (1, X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{1i}X_{2i}, X_{2i}X_{3i}, X_{1i}1\{X_{1i} > t_1\}, X_{2i}1\{X_{2i} > t_2\}, X_{1i}1\{X_{1i} > t_1\}X_{2i}1\{X_{2i} > t_2\})^\top$ where t_1 and t_2 are the sample medians of $\{X_{1i}\}_{i\in[n]}$ and $\{X_{2i}\}_{i\in[n]}$, respectively.
(viii) Lasso: the logistic probability model with regressor H_{p_n} and post-Lasso coefficient estimator $\hat{\theta}^{post}_{a,s}(\tau)$. Lasso is only applied to the case when there are four auxiliary regressors with $H_{p_n}(X_i) = (1, X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{1i}^2, X_{2i}^2, X_{3i}^2, X_{1i}X_{2i}, X_{2i}X_{3i}, X_{1i}1\{X_{1i} > t_1\}, X_{2i}1\{X_{2i} > t_2\}, X_{1i}1\{X_{1i} > t_1\}X_{2i}1\{X_{2i} > t_2\})^\top$. The post-Lasso estimator $\hat{\theta}^{post}_{a,s}(\tau)$ is defined in (A.2). The choice of tuning parameter and the estimation procedure are detailed in Section B.3.

Table 5
QTEs on total savings (four auxiliary regressors).

	NA	LP	ML	LPML	MLX	LPMLX	NP	Lasso
25%	1.105	1.473	1.105	1.105	1.105	1.105	1.105	1.105
	(0.564)	(0.564)	(0.564)	(0.564)	(0.357)	(0.319)	(0.188)	(0.564)
50%	3.682	3.682	3.682	3.682	3.682	3.682	3.682	3.682
	(1.010)	(1.033)	(0.939)	(0.939)	(0.958)	(1.033)	(0.939)	(0.939)
75%	7.363	8.100	7.363	7.363	7.363	7.363	7.363	7.363
	(3.757)	(3.757)	(3.757)	(3.569)	(3.757)	(3.663)	(3.663)	(3.757)

Notes: The table shows QTE estimates of the effect of the bank account subsidy on household total savings at quantiles 25%, 50%, and 75% when four auxiliary regressors (baseline value of total savings, household size, age, and married female dummy) are used in the regression adjustment models. Standard errors are in parentheses.

Table 6
Test for the difference between two QTEs on total savings.

	NA	LP	ML	LPML	MLX	LPMLX	NP	Lasso
50%–25%	2.577	2.209	2.577	2.577	2.577	2.577	2.577	2.577
	(0.939)	(1.104)	(0.939)	(0.939)	(0.958)	(1.033)	(0.845)	(0.911)
75%–50%	3.682	4.418	3.682	3.682	3.682	3.682	3.682	3.682
	(3.757)	(3.663)	(3.663)	(3.287)	(3.475)	(3.287)	(3.663)	(3.757)
75%–25%	6.259	6.627	6.259	6.259	6.259	6.259	6.259	6.259
	(3.851)	(3.757)	(3.757)	(3.695)	(3.588)	(3.569)	(3.287)	(3.832)

Notes: The table presents tests for the difference between two QTE estimates of the effect of the bank account subsidy on household total savings when there are four auxiliary regressors: baseline value of total savings, household size, age, and married female dummy. Standard errors are in parentheses.

This observation holds for most of the specification and estimation methods of the auxiliary regressions.[13] For example, in Table 4, the standard errors for the "LPML" QTE estimates are 16.7% less than those for the QTE estimate without adjustment at the 25th percentile, respectively. For another example in Table 5, at the 25th percentile, the standard error for the "LPMLX" QTE estimates is 43.4% less than that for the QTE estimate without adjustment. At the median, the standard error for the "LPML" QTE estimates is 7% less than that for the QTE estimate without adjustment.

Second, there is substantial heterogeneity in the impact of the subsidy on total savings. In particular, we observe larger effects as the quantile indexes increase, which is consistent with the findings in Dupas et al. (2018). For example, Table 5 shows that, although the treatment effects are all positive and significantly different from zero at quantiles 25%, 50%, and 75%, the magnitude of the effects increases by over 200% from the 25th percentile to the median and by around 100% from the median to the 75th percentile.

The second observation suggests that the heterogeneous effects of the subsidy on savings are sizable economically. To evaluate whether these effects are statistically significant, we report statistical tests for the heterogeneity of the QTEs in Table 6. Specifically, we test the null hypotheses: $H_0 : q(0.5) - q(0.25) = 0$, $H_0 : q(0.75) - q(0.5) = 0$, and $H_0 : q(0.75) - q(0.25) = 0$. Table 6 shows that only the difference between the 50th and 25th QTEs is statistically significant at the 5% significance level.

How does the variation in the impact of the subsidy appear across the distribution of total savings? The QTEs on the distribution of savings are plotted in Fig. 1, where the shaded areas represent the 95% confidence region. The figure shows that the QTEs seem insignificantly different from zero below about the 20th percentile. At higher levels to near the 80th percentile, the treatment group savings exceed the control group savings at an accelerated rate, yielding increasingly significantly positive QTEs. Beyond the 80th percentile, the QTEs again become insignificantly different from zero. These findings point to notable distributional heterogeneity in the impact of the subsidy on savings.

8. Conclusion

This paper proposes the use of auxiliary regression to incorporate additional covariates into estimation and inference relating to unconditional QTEs under CARs. The auxiliary regression model may be estimated parametrically, nonparametrically, or via regularization if there are high-dimensional covariates. Both estimation and bootstrap inference methods are robust to potential misspecification of the auxiliary model and do not suffer from the conservatism due to the CAR. It is shown that efficiency can be improved when including extra covariates. When the auxiliary regression is correctly specified, the regression-adjusted estimator further achieves the minimum asymptotic variance. In both the simulations and the empirical application, the proposed regression-adjusted QTE estimator performs well. These results and the

13 The efficiency gain from the "NP" adjustment is not the only reason for its small standard error at the 25% QTE. Another reason is that the treated outcomes around this percentile themselves do not have much variation.

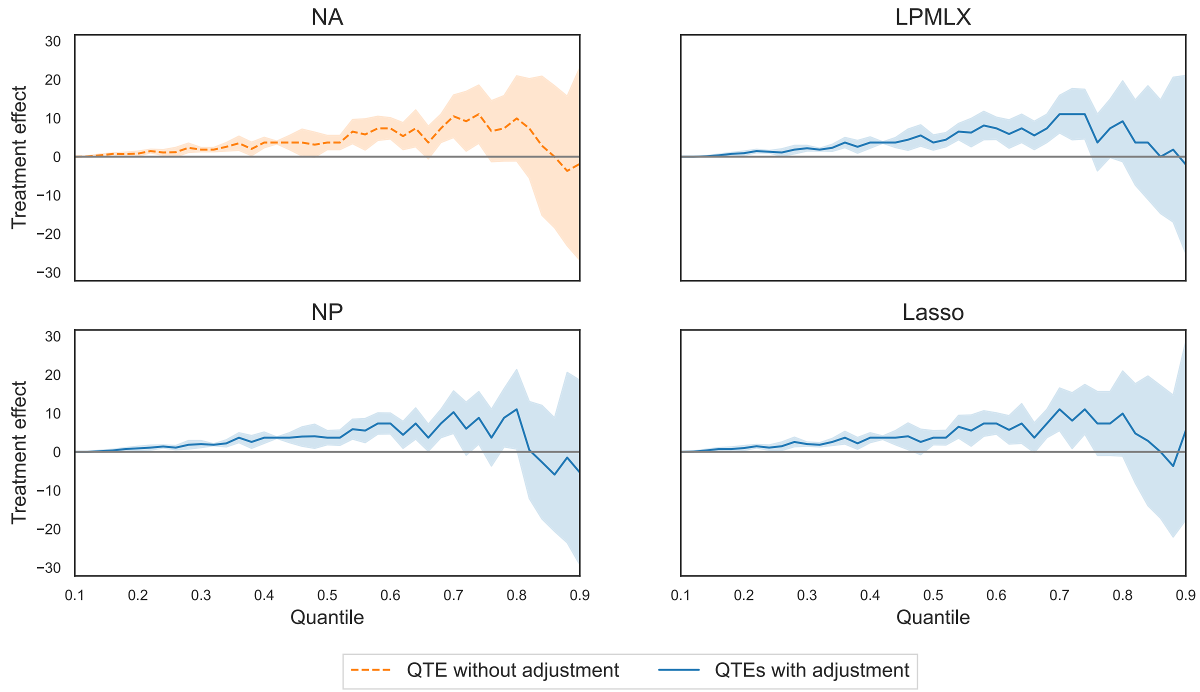

Fig. 1. Quantile treatment effects on the distribution of total savings.
Notes: The graphs in each panel of the figure plot the QTE estimates of the effect of the bank account subsidy on the distribution of household total savings when there are four auxiliary regressors: baseline value of total savings, household size, age, and married female dummy. The shadowed areas display 95% confidence regions.

robustness of the methods to auxiliary model misspecification reflect the aphorism widespread in scientific modeling that all models may be wrong, but some are useful.[14]

Acknowledgments

We thank the Managing Editor, Elie Tamer, the Associate Editor and three anonymous referees for many useful comments that helped to improve this paper. We are also grateful to Michael Qingliang Fan and seminar participants from the 2022 Econometric Society Australasian Meeting, the 2021 Nanyang Econometrics Workshop, University of California, Irvine, and University of Sydney for their comments.

Funding

Yichong Zhang acknowledges financial support from the Singapore Ministry of Education under Tier 2 grant No. MOE2018-T2-2-169, the NSFC, China under the grant No. 72133002, and a Lee Kong Chian fellowship. Peter C. B. Phillips acknowledges support from NSF Grant No. SES 18-50860, a Kelly Fellowship at the University of Auckland, New Zealand, and a Lee Kong Chian Fellowship. Yubo Tao acknowledges the financial support from the Start-up Research Grant of University of Macau (SRG2022-00016-FSS). Liang Jiang acknowledges support from MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 18YJC790063).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2022.08.010.

14 The aphorism "all models are wrong, but some are useful" is often attributed to the statistician George Box (1976). But the notion has many antecedents, including a particularly apposite remark made in 1947 by John von Neumann (2019) in an essay on the empirical origins of mathematical ideas to the effect that "truth ... is much too complicated to allow anything but approximations".

References

Abadie, A., Chingos, M.M., West, M.R., 2018. Endogenous stratification in randomized experiments. Rev. Econ. Stat. 100 (4), 567–580.
Anderson, S.J., McKenzie, D., 2021. Improving business practices and the boundary of the entrepreneur: a randomized experiment comparing training, consulting, insourcing and outsourcing. J. Polit. Econ. 130 (1), 157–209.
Ansel, J., Hong, H., Li, J., 2018. OLS and 2SLS in randomised and conditionally randomized experiments. J. Econ. Stat. 238, 243–293.
Athey, S., Imbens, G.W., 2017. The econometrics of randomized experiments. In: Handbook of Economic Field Experiments. Vol. 1. Elsevier pp. 73–140.
Bai, Y., 2022. Optimality of matched-pair designs in randomized controlled trials. Amer. Econ. Rev. (forthcoming).
Bai, Y., Shaikh, A., Romano, J.P., 2021. Inference in experiments with matched pairs. J. Amer. Statist. Assoc. (forthcoming).
Banerjee, A., Duflo, E., Glennerster, R., Kinnan, C., 2015. The miracle of microfinance? Evidence from a randomized evaluation. Am. Econ. J. Appl. Econ. 7 (1), 22–53.
Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C., 2017. Program evaluation with high-dimensional data. Econometrica 85 (1), 233–298.
Bitler, M.P., Gelbach, J.B., Hoynes, H.W., 2006. What mean impacts miss: distributional effects of welfare reform experiments. Amer. Econ. Rev. 96 (4), 988–1012.
Bloniarz, A., Liu, H., Zhang, C.-H., Sekhon, J.S., Yu, B., 2016. Lasso adjustments of treatment effect estimates in randomized experiments. Proc. Natl. Acad. Sci. 113 (27), 7383–7390.
Box, G.E., 1976. Science and statistics. J. Amer. Statist. Assoc. 71 (356), 791–799.
Bugni, F.A., Canay, I.A., Shaikh, A.M., 2018. Inference under covariate-adaptive randomization. J. Amer. Statist. Assoc. 113 (524), 1741–1768.
Bugni, F.A., Canay, I.A., Shaikh, A.M., 2019. Inference under covariate-adaptive randomization with multiple treatments. Quant. Econ. 10 (4), 1747–1785.
Bugni, F.A., Gao, M., 2021. Inference under covariate-adaptive randomization with imperfect compliance. arXiv Preprint arXiv:2102.03937.
Burchardi, K.B., Gulesci, S., Lerva, B., Sulaiman, M., 2019. Moral Hazard: experimental evidence from tenancy contracts. Q. J. Econ. 134 (1), 281–347.
Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H.C., McKenzie, D., Mensmann, M., 2017. Teaching personal initiative beats traditional training in boosting small business in West Africa. Science 357 (6357), 1287–1290.
Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. Econometrica 81 (6), 2205–2268.
Chong, A., Cohen, I., Field, E., Nakasone, E., Torero, M., 2016. Iron deficiency and schooling attainment in Peru. Am. Econ. J. Appl. Econ. 8 (4), 222–255.
Cohen, P.L., Fogarty, C.B., 2020. No-harm calibration for generalized oaxaca-blinder estimators. arXiv preprint arXiv:2012.09246.
Crépon, B., Devoto, F., Duflo, E., Parienté, W., 2015. Estimating the impact of microcredit on those who take it up: evidence from a randomized experiment in Morocco. Am. Econ. J. Appl. Econ. 7 (1), 123–150.
Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2013. Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. Q. J. Econ. 128 (4), 1499–1545.
Dupas, P., Karlan, D., Robinson, J., Ubfal, D., 2018. Banking the unbanked? Evidence from three countries. Am. Econ. J. Appl. Econ. 10 (2), 257–297.
Firpo, S., 2007. Efficient semiparametric estimation of quantile treatment effects. Econometrica 75 (1), 259–276.
Fogarty, C.B., 2018. Regression-assisted inference for the average treatment effect in paired experiments. Biometrika 105 (4), 994–1000.
Freedman, D.A., 2008a. On regression adjustments in experiments with several treatments. Ann. Appl. Stat. 2 (1), 176–196.
Freedman, D.A., 2008b. On regression adjustments to experimental data. Adv. Appl. Math. 40 (2), 180–193.
Greaney, B.P., Kaboski, J.P., Van Leemput, E., 2016. Can self-help groups really be "self-help"? Rev. Econom. Stud. 83 (4), 1614–1644.
Hahn, J., Hirano, K., Karlan, D., 2011. Adaptive experimental design using the propensity score. J. Bus. Econom. Statist. 29 (1), 96–108.
Hahn, J., Liao, Z., 2021. Bootstrap standard error estimates and inference. Econometrica 89 (4), 1963–1977.
Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71 (4), 1161–1189.
Hu, Y., Hu, F., 2012. Asymptotic properties of covariate-adaptive randomization. Ann. Statist. 40 (3), 1794–1815.
Jakiela, P., Ozier, O., 2016. Does Africa need a Rotten Kin theorem? Experimental evidence from village economies. Rev. Econom. Stud. 83 (1), 231–268.
Jiang, L., Liu, X., Phillips, P.C.B., Zhang, Y., 2021. Bootstrap inference for quantile treatment effects in randomized experiments with matched pairs. Rev. Econ. Stat. (forthcoming).
Kallus, N., Mao, X., Uehara, M., 2020. Localized debiased machine learning: efficient inference on quantile treatment effects and beyond. arXiv preprint arXiv:1912.12945.
Karlan, D., Ratan, A.L., Zinman, J., 2014. Savings by and for the poor: A research review and agenda. Rev. Income Wealth 60 (1), 36–78.
Lei, L., Ding, P., 2021. Regression adjustment in completely randomized experiments with a diverging number of covariates. Biometrika 108 (4), 815–828.
Li, X., Ding, P., 2020. Rerandomization and regression adjustment. J. R. Stat. Soc. Ser. B Stat. Methodol. 82 (1), 241–268.
Lin, W., 2013. Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique. Ann. Appl. Stat. 7 (1), 295–318.
Liu, H., Tu, F., Ma, W., 2020. A general theory of regression adjustment for covariate-adaptive randomization: OLS, Lasso, and beyond. arXiv preprint arXiv:2011.09734.
Liu, H., Yang, Y., 2020. Regression-adjusted average treatment effect estimates in stratified randomized experiments. Biometrika 107 (4), 935–948.
Lu, J., 2016. Covariate adjustment in randomization-based causal inference for 2K factorial designs. Statist. Probab. Lett. 119, 11–20.
Ma, W., Hu, F., Zhang, L., 2015. Testing hypotheses of covariate-adaptive randomized clinical trials. J. Amer. Statist. Assoc. 110 (510), 669–680.
Ma, W., Qin, Y., Li, Y., Hu, F., 2020. Statistical inference for covariate-adaptive randomization procedures. J. Amer. Statist. Assoc. 115 (531), 1488–1497.
Muralidharan, K., Sundararaman, V., 2011. Teacher performance pay: experimental evidence from India. J. Polit. Econ. 119 (1), 39–77.
Negi, A., Wooldridge, J.M., 2020. Revisiting regression adjustment in experiments with heterogeneous treatment effects. Econometric Rev. 40 (5), 1–31.
Olivares, M., 2021. Robust Permutation Test for Equality of Distributions under Covariate-Adaptive Randomization. Working Paper, University of Illinois at Urbana Champaign.
Shao, J., Yu, X., 2013. Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. Biometrics 69 (4), 960–969.
Shao, J., Yu, X., Zhong, B., 2010. A theory for testing hypotheses under covariate-adaptive randomization. Biometrika 97 (2), 347–360.
Tabord-Meehan, M., 2021. Stratification trees for adaptive randomization in randomized controlled trials. arXiv preprint arXiv:1806.05127.
van der Vaart, A., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer, New York.
von Neumann, J., 2019. The mathematician. In: Mathematics: People, Problems, Results, second ed. Chapman and Hall/CRC.
Wei, L., 1978. An application of an urn model to the design of sequential controlled clinical trials. J. Amer. Statist. Assoc. 73 (363), 559–563.
Ye, T., 2018. Testing hypotheses under covariate-adaptive randomisation and additive models. Stat. Theory Rel. Fields 2 (1), 96–101.
Ye, T., Shao, J., 2020. Robust tests for treatment effect in survival analysis under covariate-adaptive randomization. J. R. Stat. Soc. Ser. B Stat. Methodol. 82 (5), 1301–1323.
Ye, T., Yi, Y., Shao, J., 2022. Inference on average treatment effect under minimization and other covariate-adaptive randomization methods. Biometrika 109 (1), 33–47.
Zhang, Y., Zheng, X., 2020. Quantile treatment effects and bootstrap inference under covariate-adaptive randomization. Quant. Econ. 11 (3), 957–982.
Zhao, A., Ding, P., 2021. Covariate-adjusted Fisher randomization tests for the average treatment effect. J. Econometrics 225 (2), 278–294.