Chapter 2: The effects of type II diabetes on financial distress and labor supply: evidence from administrative data

2.1 Introduction

Diabetes is one of the most diffuse diseases in the United States today, affecting nearly 15% of American adults in 2020. Its prevalence has increased five percentage points since 2001 (Centers of Disease Control and Prevention, 2021). Its severity makes it one of the top ten causes of death amongst Americans (Xu et al., 2022), and a growing body of research shows an association between type II diabetes and all cause mortality (Raghavan et al., 2019; Yang et al., 2019). Diabetes has been called the "biggest epidemic of the twenty-first century" (Tabish, 2007) and has prompted national programs[11] and international collaboration[12] to reduce its prevalence. And while the severity of the disease is almost universally recognized in terms of its physical effects, its cost to society is perhaps less appreciated but may be equally grave.

Data from the American Diabetes Association (ADA) attributed $214 billion of total direct medical costs in 2017, and, of that, over $100 billion for Medicaid and Medicare recipients alone (Shrestha et al., 2018).[13] Moreover, the 2013 National Health

[11] https://www.cdc.gov/diabetes/prevention/index.html

[12] https://www.who.int/initiatives/the-who-global-diabetes-compact

[13] Indirect costs, like reduced productivity and absenteeism, accounted for over $250 billion in 2017.

Interview Survey reported that half of adults living with diabetes experience financial stress, and that those with diabetes were twice as likely to report cost related non-adherence (CRN) due to cost of disease management compared to those not living with diabetes (Patel et al., 2016). More recently, novel metrics have been used to better understand the financial distress of diabetic patients, and have further demonstrated a positive association between disease severity and distress (Patel et al., 2022).

An over-representation of diabetes in low-income and socially disadvantaged groups prompts a discussion of the possibility that chronic illness is, at least in part, socially determined. Coughlin et al. (2021) find that the individuals living with five co-morbidities are about five-times more likely to have annual incomes below $15,000 as those with zero morbidities. This highlights the difficulty in assessing the impact of a disease on financial distress, namely that reverse causality is a plausible alternative explanation. Does diabetes cause financial distress, or does chronic stress (contributed in part by financial strain) bring about diabetes? For a physician tasked with managing the symptoms of the disease, this distinction is maybe less important, but for policymakers and public health officials, the distinction highlights dramatically different sets of policies. In this essay, I use electronic medical records of individuals receiving blood tests to identify the onset of diabetes, and a panel of consumer credit records from Experian Credit Bureau to straighten out the possibility of reverse causation that has typically plagued the literature. I expand on recent literature that finds that marginal diabetes diagnosis adds costs with little health benefit (Alalouf et al., 2024). More broadly, this essay examines the financial effects of diabetes diagnosis by examining changes in consumption (through borrowing), default on debt payments, and labor supply.

In addition to measuring the financial consequences of diabetes on the expense side of the balance sheet, this essay also examines labor responses to diabetes diagnosis. Labor supply must be viewed as a complementary outcome in order to understand the scope of economic impacts. It has long been recognized that one financial coping mechanism to financial burden is labor supply. The economics literature conceptualizing and analyzing this phenomenon has traditionally examined the ability of the household, or at least another member of the household, to moderate labor supply in response to financial shock (Lundberg, 1985), but the relationship is captured in the neoclassical models of individual labor supply too. In these models, the worker chooses how much labor to supply in order to maximize utility subject to a budget constraint. Working more allows the individual to consume more, but prevents the individual from spending more time in leisure. Hence, the optimal choice depends on the individual's preferences for consumption and leisure. However, medical disease adds a wrinkle to this decision, as the worker may be unable to perfectly adjust his labor supply. In particular, the worker afflicted with a medical disease like diabetes may be unable to work (either due to physical limitations or added time constraints), and thus the disease may exacerbate the financial burden.

Indeed, Jowsey et al. (2012) reviewed twenty-two articles published on the time costs of health related illness, and offers the consensus that patients may spend about 2-hours per day managing their disease, with diabetic patients incurring more time costs than other chronic illnesses. Breton et al. (2013) review the literature on the ability to work of individuals living with diabetes and note that those with diabetes miss almost twice as many days per year compared to those without diabetes (absenteeism), have more than twice as much lost productivity per year, and retire about one year earlier than those without the disease. In addition, men with diabetes are more than three times more likely to stop working outside the home due to illness compared to men without diabetes, and women living with diabetes are about four-times more likely to stop working outside the home due to illness.

Yet the same difficulties of endogeneity has made causal interpretations of diabetes on labor supply difficult, too. Does diabetes alter labor supply choices, or are labor supply choices and diabetes simultaneous results of something else? Does having a chronic illness cause the patient to enter the labor force to pay for the added medical expenses (or leave the labor force due to added time costs), or do lower income and riskier jobs bring about a lifestyle that leads to diabetes? Pedron et al. (2019) conduct a systematic review of nearly 4,000 articles seemingly-related to labor market outcomes of diabetics. Recognizing the difficulty of identifying a causal effect due to the endogeneity of labor supply and disease, the authors note that not a single article was of high enough quality to definitively answer questions related to labor force participation. Amongst the reasons why the authors concluded this, the overarching theme is the lack of administrative data either (and often in both) the medical and labor data. Still, the authors conclude that, though the estimates are often dramatically inconsistent between studies, there is typically a negative association between diabetes and employment.

Using a rare linkage between electronic medical records and state employment records, I aim to fill a void in the literature, noted by Pedron et al. (2019), to causally assess the impact of diabetes diagnosis on labor market outcomes.

In this chapter, I begin the analysis with event study difference in differences models to assess the extent to which consumption, debt repayment, and labor market outcomes are affected by the onset of diabetes. While this empirical strategy has some important assumptions that must be met, most notably that differences in outcomes between treated and untreated groups are related to the treatment and not to idiosyncratic characteristics of the treated group, if done well it allows us to capture the average treatment effect (ATE) of diagnosis. The ATE is intuitive answer to the question, "What effect does diabetes diagnosis have on financial distress, and labor supply?"

I also implement a regression discontinuity design (RDD) to further isolate the causal effects of being diagnosed with diabetes. I do this by exploiting the diagnosis recommendations of the ADA, namely a strict cutoff in the level of blood hemoglobin A1C (HbA1C) of 6.5%. The intuition of this approach is that it compares behaviors of individuals who are nearly the same in terms of their disease severity but different in their diagnosis status. It limits the scope of generalization to individuals right on the margin, giving an estimate of the Local ATE (LATE), but it purports to even further disentangle any form of reverse causality because it relies on the fact that individuals cannot strategically place themselves on either side of the diagnosis threshold. In a sense, this second method also allows me to distinguish between the effect of diagnosis on economic outcomes rather than of disease progression itself.

Using this same approach, Alalouf et al. (2024) find that patients diagnosed with diabetes at the margin incurred $1,000 of increased annual medical costs for up to six years after diagnosis with little or no evidence of improved clinical measures of health in subsequent years compared to marginally non-diagnosed patients. While this sum of money may seem negligible, it is important to note that surveys have traditionally shown that nearly half of individuals have little or no precautionary savings (Lusardi, 2011; Lusardi et al., 2011; Beshears et al., 2018), and that 40% of individuals would be unable to withstand an unexpected expense of $400 (Chen et al., 2019). Indeed, Chetty and Szeidl (2007) show that the typical household's monthly expenditures are not very flexible, with only about 50% of the monthly expenditures being uncommitted (though this classification includes necessities like food, utilities, ecc.). In a situation which is absent of additional time costs and health limitations, the constrained ability to adjust consumption paired with added financial burden suggest the possibility that labor could adjust to mitigate the burden. It is clear, however, that these conditions are unlikely to be met in the context of medical disease, and so empirical exploration is necessary.

In this essay, I aim to contribute to this conversation by examining the economic consequences of chronic disease diagnosis, which adds both to the literature particularly about type II diabetes cited above and to the literature that more broadly matches health records and financial administrative data to assess the financial consequences of medical disease, sometimes called financial toxicity (Dobkin et al., 2018; Gupta et al., 2018; Ramsey et al., 2013; Scott et al., 2022; Shankaran et al., 2022). I match electronic medical records from the Ohio State University Wexner Medical Center for individuals who received HbA1C tests between 2017 and 2021 with consumer credit records from Experian Credit Bureau, and with individual labor force data from the Ohio Department of Job and Family Services to examine differences in financial distress and labor supply in the population of diagnosed diabetic patients.

Using the event study difference in differences models, I find minor evidence of reduced consumption that results from diabetes diagnosis. The exception to this is that I find a sustained decrease in the availability of credit that patients have after diagnosis. While changes in consumption appear minor, I find evidence of increased delinquency in the two years after diagnosis, suggesting that patients might be substituting costly future consumption for immediate consumption. I find evidence that this decrease in the likelihood of debt payments also impacts credit score, which is a major determinant of the cost of future credit. In particular, I find a precisely estimated and sustained decrease of about 2-4 points to a consumer's credit score. Consumers are no more likely to have below prime credit scores, but I show evidence that the cost of future credit may increase. A decrease of about 6% in the probability of having a super prime credit score suggests that individuals can likely still access credit, though at an increased cost than pre-diagnosis. The effects noted in aggregate are driven by older adults and patients living alone. Whereas pressure to increase labor supply may be expected in contexts where financial burden is not met with time costs, I find that employment declines 13% relative to the baseline immediately following diagnosis, and remains decreased for at least two years. This decrease in labor supply may exacerbate the financial stress already felt by newly-diagnosed patients.

Results from the regression discontinuity design detail a slightly different story. While the event study difference in differences finds that patients are typically maintaining consumption and increasing delinquency, for patients marginally diagnosed, I find evidence that consumption declines relative to the marginally non-diagnosed, and I find little evidence of decreased debt repayment. Likewise, I find no difference in credit scores. I also find no differences in labor outcomes between these groups. The results from the regression discontinuity suggest that for patients at the margin of diagnosis, the added financial and time burdens of diabetes may be managed simply through changes in contemporaneous consumption.

The rest of the essay is as follows. In the next section, I begin by examining the budget constraint of individuals to better formalize which economic consequences might reasonably be the result of increased financial obligations through medical disease. I then define the constructs for this analysis as measured in consumer credit records and labor data. After describing the empirical methods, I discuss the results.

2.2 Conceptual motivation

Financial toxicity is a topic of study originally imported from medical research that is finding popularity in the health economics literature. The core of this concept is that medical disease has predictable physical *and* financial effects. When a patient is diagnosed or treated with a chronic medical condition, numerous changes to the patient's life and lifestyle follow. Perhaps the most immediate thing that the patient and loved ones might notice are physical side effects. A patient, for example, treated for cancer may lose his hair or become weak and frail. A patient diagnosed with diabetes is more substantially likely to become blind and to lose a limb relative to the population at large. In fact, it is estimated that individuals with type II diabetes may have a life expectancy of at about 15 years less than someone without diabetes (Kaptoge et al., 2023). And while these physical effects might be the most dramatic and perhaps even the most concerning, typically the effects of medical disease unfortunately do not stop there. Indeed, the financial effects of medical disease can often be just as devastating, though the effect often extends beyond just the patient to the household as a whole, and even sometimes across multiple generations.

This consideration of the economic consequences of medical disease is increasingly called financial toxicity (Shankaran et al., 2022; Knight et al., 2018; Santacroce and Kneipp, 2019; Ramsey et al., 2013). The literature connecting medical shocks to financial distress has shown increased rates of bankruptcy (Dobkin et al., 2018; Ramsey et al., 2013), non-payment and adverse financial events (Shankaran et al., 2022), and medical non-compliance (Knight et al., 2018; Moulton et al., 2022). Just as a patient might lose his strength from a prescribed medication, the patient's financial health may also be jeopardized from costly treatments and increased time demands that may pull him away from his typical work schedule. Financial toxicity is an attractive concept because it is intuitive: it treats financial distress as an often-overlooked symptom of medical disease. But while this concept is intuitive on its face, exploring the motivation more rigorously has some benefit. In particular, it reveals why financial distress is a reasonable expectation from changes in financial obligations, but it also gives insight into other economic consequences that may be reasonably expected.

Consider a utility-maximizing individual, where the individual's utility is a function increasing in consumption (x) and health (H), $U = U(x, H)$. An individual maximizes U subject to a budget constraint. In particular, the consumer's consumption is constrained by his wage income, y, borrowed financial capital, b, and net assets, $A - D$, such that his period-specific budget constraint is

$$y^t + b^t + A_{t-1}(1 + r_A) - D_{t-1}r_D \geq m + p_x x^t$$

where

$$(1)\ D_t = b_t + D_{t-1}(1 + r_D)$$

$$(2)\ A_t = A_{t-1}(1 + r_A) + s_t$$

$$(3)\ s_t = y - p_x x - m + b_t$$

In this budget constraint, r_D is the rate at which the consumer borrows capital in the credit market at time t, and is thus largely endogenous. By contrast, the rate r_A is the rate at which assets grow over time, and is largely exogenous to the consumer. When patients incur increased medical costs, m, several behavioral changes could occur to offset this added financial burden in the budget constraint.

Perhaps most simply, the patient could decrease consumption, x. As mentioned above, Chetty and Szeidl (2007) use the Consumer Expenditure Survey to show that the modern realities of monthly mortgage or rent payments, auto and student loan payments, and insurance premiums constrain an individual's ability to moderate consumption dramatically. Accordingly, this is an area where credit may especially be useful for diagnosed patients. While a patient may not be able to moderate his monthly consumption much, he may be able to increase credit utilization, that is to consume closer to the limit of his available credit. An extension of this concept is that an individual may increase his available credit by opening a new line of revolving credit (e.g. credit card). With these considerations in mind, it is also worth noting that the basic notion of consumption is notoriously challenging to measure in consumer credit data, since credit balances are monthly snapshots of a credit line, and thus conflate stocks and flows. Credit records also miss the observance of consumption through cash or checking accounts.[14] Researchers who use consumer credit data have thus considered attributes of an individual's credit file like balance to credit ratio, and the presence of new trades (e.g. credit cards, auto loans, mortgages, and personal loans) as ways to infer discretionary consumption from credit records. The consumer credit data that I link to in this analysis allows me to examine these outcomes.

A second possibility is that individuals may respond to increased medical costs by altering their labor supply. Absent physical or time constraints, the relationship is clear: all else equal, increased medical costs increases labor supply. But while at the margin, this is certainly plausible, with chronic diseases in particular the relationship is not quite as simple. At least two characteristics of chronic disease give credence to the inverse relationship: first, chronic disease can be physically debilitating, and hence a patient may be unable to increase his labor supply (or, instead, may even reduce it). Second, management of chronic illness can be extremely time consuming, and so the actual time constraints of a week may also contribute to a reduction in labor supply. Indeed, descriptive research on diabetic patients has generally shown a negative association with labor force attachment and diabetes (Kahn, 1998; Minor and MacEwan, 2016; Tunceli et al., 2005; Ng et al., 2001; Pedron et al., 2019; Breton et al., 2013).

Individuals with severely unmanaged diabetes may have decreased ability to engage with the labor force, providing justification for the negative association between labor force attachment and diabetes status, but those at the margin should have largely the same ability to adjust labor supply as desired. One caveat to this latter assertion,

[14] The prominence of cash payments is historically decreasing and is estimated to account for only about 6-7% of the number of transactions and 20% of the value of all purchases (Cubides and O'Brien, 2022).

however, one which further bolsters the negative association between diabetes and labor force participation, may be that marginal diagnosis results in increased time costs. Indeed, Jowsey et al. (2012) reviewed twenty-two articles published on the time costs of health related illness, and offers the consensus that patients may spend about 2-hours per day managing their disease, with diabetic patients incurring more time costs than other chronic illnesses. To the extent that marginal diagnosis of diabetes incurs added time costs, the pressure to increase labor supply may be constrained. Using a rare linkage between patient records, credit records, and administrative labor force data, I am able to explore the possibility that increased medical costs result in changes in labor supply. In particular, I examine changes in quarterly labor force participation, quarterly wage earnings, employer count, and weeks worked.[15] What I observe is wage earnings derived from engagement in the labor market, and so this will miss non-labor income. While this measurement is imprecise, for those in the labor force, wage earnings has traditionally comprised the major part of personal income in the United States (Lawson et al., 2014).

Finally, a person incurring added financial strain may simply consume from the future by neglecting his period-specific financial obligations, D_t. Practically, he might do this through failure to make debt payments[16] and/or consuming from a greater share of the credit that is available to him. An equivocal way of thinking about defaulting on debt is to consider it as a costly form of borrowing from the future (Braxton et al.,

[15] A major caveat of observing weeks worked is that the variable collected is the maximum number of weeks worked in a quarter, and thus conceals the possibility of switching jobs, adding a second job, or changing the number of hours worked per week.

[16] For example, delinquency (i.e. being past due on payments to creditors), chargeoffs (i.e. the credit has attempted and failed to collect financial obligations, and so has closed the account) collections, bankruptcy, and foreclosure.

2020). In the United States, a person seeking credit is evaluated on how likely he is to repay the debt, typically referred to as a credit score. Credit score is a signal to lenders about the credit worthiness of borrowers, and is used as a major factor in determining the cost of future borrowing.

There are several scoring algorithms but each purports to use similar measures of similar weight, and ideally gives consistent information to lenders about the credit worthiness of potential borrowers. While the data that I use provides the Vantage Score, the Fair Isaac Corporation (FICO) score is the other widely-used score, which also publishes the relative weight of the factors in a credit score.[17] The most important factors in the credit score are payment history (35%) and credit utilization (30%). Hence, in finding tangible outcomes that indicate inter-temporal substitution, credit score may be one aggregate measure that I examine. Credit score not only provides an aggregate measure that is partially driven by payment history, but also of credit access. I consider the extent to which a change in credit score might be impactful in credit access by considering changes in the probability of being below prime (less than 620), of being deep prime (below 580), and of being super prime (above 720).

These classifications of credit score are widely recognized groupings of individuals, and may better inform the intensity of any change in credit score. These classifications also serve as buckets in which to think of the cost of credit. For classes of low credit score (e.g. deep subprime credit scores), consumers might be credit constrained, or face costs that make accessing credit unlikely; whereas those with super prime credit scores, are both likely to be approved for a new line of credit likely to face a cost of borrowing that is often significantly lower than consumers with lower credit scores.

[17] https://www.myfico.com/credit-education/whats-in-your-credit-score

I measure outcomes related to failure to repay debt, and I measure credit score. Finding tangible representations of concepts is imperfect, and so while there might be a degree to which these selected attributes overlap, my aim is to identify reasonable areas that we might expect to see financial difficulty to manifest.

2.2.1 Heterogeneity

The budget constraint as I have written it obfuscates heterogeneity in the effect of diabetes diagnosis on economic outcomes. In particular, the budget constraint simply described medical costs as m. But in reality, especially in the American context of healthcare, m is really a (sometimes complex) function that relates the charges of the procedures with the costs that actually face the patient. Dobkin et al. (2018) model this as $(1 - \lambda_m)m$, which seems a reasonable simplification. For the purposes of this essay, the importance is that it defines differences in m by particular groups; perhaps most easily observed are the cost differences between age groups who have access to Medicare (i.e. 65+) and those who do not. Those who have federal health insurance in the form of Medicare face dramatically reduced costs than those with private insurance, or equivalently a much larger λ_m.

Past research has shown that economic hardship for the elderly on Medicare is dramatically reduced compared to non-Medicare recipients (Dobkin et al., 2018). Conversely, older adults are typically less financially savvy (Finke et al., 2017; Allgood and Walstad, 2013), consume from a fixed income, and have a high pre-disposition to co-morbitities (Kim et al., 2018; Iglay et al., 2016), which provide some justification for why the effect could be larger for older adults. I explore heterogeneity in consumption, debt repayment, and labor supply by differences in λ_m in later portions of this essay. Younger diabetic patients may also respond in more pronounced ways in the labor market. In particular, while the pressure to work is, all else equal, greater for younger adults who incur higher medical costs, a greater share of younger adults are employed in the labor market, and so the response – either to work more or to work less – may be more pronounced than it may be for older adults, many of whom may be already retired and receiving supplemental income.

I also make use of a unique characteristic of the credit panel to consider two additional types of heterogeneity. First, I consider differences between patients who were living alone at diagnosis and those who were not. This difference in household type could be the source of variation in income, savings, or access to credit. Financial burden could be greater for single individuals, as the high time costs of diabetes is borne by the individual himself and not by other household members (Lundberg, 1985). In particular, a single patient with high time cost is one who may supply less labor, substituting that time to caring for his disease, or may require medical assistance, which can be costly. I empirically examine heterogeneity in the outcome by household type of the patient at diagnosis.

Second, I explore differences in outcomes by individuals whose credit use is high relative to their credit availability ($> 30\%$) in the period prior to diagnosis compared to individuals with lower credit utilization. 30% is a utilization rule that is often recommended by financial planning professionals. Individuals with a utilization rate above 30% are defined to be credit constrained.

2.3 Data

The dataset used in this analysis combines patient electronic medical record data (EMR), cost, and insurance data from the Ohio State University Wexner Medical Center with consumer credit data from the Ohio Consumer Credit Panel (Ohio CCP), sourced from Experian Credit Bureau,[18] and a random 20% sample of labor supply data from the Ohio Department of Job and Family Services (ODJFS).[19] The patient data includes the universe of patients tested for HbA1C levels at Ohio State University medical facilities between 2017 and 2021. Using a secure process that matches the hashed social security number of the patient, I integrate the consumer credit records from the Ohio CCP. This dataset is comprised of the universe of consumer credit records from the state of Ohio quarterly from Q4 2017 to Q4 2021, approximately 8.8 million individuals or about 95% of the Ohio adult population.[20] It allows me to see a wide array of important financial characteristics at the quarterly level such as debt levels, measures of financial delinquency, public filings like bankruptcies, and credit score. Additionally, because I have the universe of consumers in Ohio, I am

[18] There is a growing number of studies that use data from credit bureaus linked to medical records to understand the financial consequences of disease, e.g. (Dobkin et al., 2018; Gupta et al., 2018; Shankaran et al., 2022; Scott et al., 2022; Carlton et al., 2023).

[19] This data is quarterly data on all employed individuals in the state of Ohio, and includes quarterly wages and weeks employed. It is compiled as part of the Ohio Longitudinal Data Archive (OLDA), which is a project of the Ohio Education Research Center (oerc.osu.edu) and provides researchers with centralized access to administrative data. The OLDA is managed by The Ohio State University's Center for Human Resource Research (chrr.osu.edu) in collaboration with Ohio's state workforce and education agencies (ohioanalytics.gov), with those agencies providing oversight and funding. For information on OLDA sponsors, see http://chrr.osu.edu/projects/ohio-longitudinal-data-archive .

[20] Prior studies estimate that approximately 11 percent of adults in the U.S. do not have a credit file (Brevoort et al., 2016). However, coverage in credit data has expanded over the past few years. Further, Ohio has a very small immigrant population and thus fewer people who have not yet established a credit file compared to states like California, Texas, Florida, and New York.

also able to identify household members of the patients, and thus to create measures of household type for patients.

2.3.1 Patient sample

I classify patients as diagnosed with diabetes when their interaction with the medical system in the form of an A1C test results in either a diabetes-related ICD-10 diagnosis code, or a diabetes-related medication. To be classified as newly-diabetic, I require that the first test during the study period 2017-2021 result in a non-diagnosis classification (i.e. no ICD-10 code and no diabetes medication), and that a subsequent test result in a diagnosis classification. In the event study difference in differences models, I consider diagnosis to occur in the quarter when the first ICD-10 code or diabetes medication is observed.

By contrast, in order to compare results to and expand on the work of Alalouf et al. (2024), in the fuzzy RDD approach I allow diagnosis to occur within four quarters of the HbA1C test, which is a conservative approach to disentangle any delayed reporting between the test and the medical records. This approach is obviously more "test-centric" compared to the event study difference in differences models, which is agnostic to the test. While there is potential for recording error (for example, if the doctor did not record a diabetes-related ICD-10 code for a non-diabetes related interaction with the health system), medical systems are financially reimbursed in large part due to the complexity of the patient. Thus, misreporting is unlikely to be the case. The patient population in this sample includes all individuals who were tested for HbA1C at Ohio State University medical facilities. Because Ohio State self-insures, many individuals receive a yearly blood screening, of which A1C is a component. As a result, many of these individuals are receiving regular blood tests, and not merely in response to a health crisis.

I exclude patients whose first interaction with the medical system resulted in a diagnosis for diabetes. This is to distinguish between HbA1C testing that is used for diagnosis rather than disease management. This has the effect of omitting some patients who are truly newly-diagnosed at their first interaction with the health system, but if financial distress is truly driven by diabetes, then the results should be understated. I also exclude patients who were tested while pregnant, as gestational diabetes is typically temporary. To protect patient security, and to match the level of aggregation of the administrative data, the data is aggregated to the quarter. The resulting analytic sample thus includes all the quarters of HbA1C tests for individuals who are never diagnosed with diabetes in the sample period (105,164 unique patients), and for individuals who were previously not diabetic but later become diabetic (5,982 unique patients). The resulting sample is a maximum of 111,146 unique patients.

Figure 2.1 shows the distribution of HbA1C values across the entire sample of OS-UWMC patients (left) and in the resulting analytic sample, as described above (right). As expected, because I drop patients who are tested for management rather than diagnosis, the distribution of the analytic sample has much less weight to the right of the ADA-recommended threshold.

In Table 2.1, I report summary statistics that compare patients across the OSU sample for this analysis, and – where possible – what is published in Alalouf et al. (2024), which is the closest article to this essay, and which provides data on patients from two other sources, described in greater detail below.

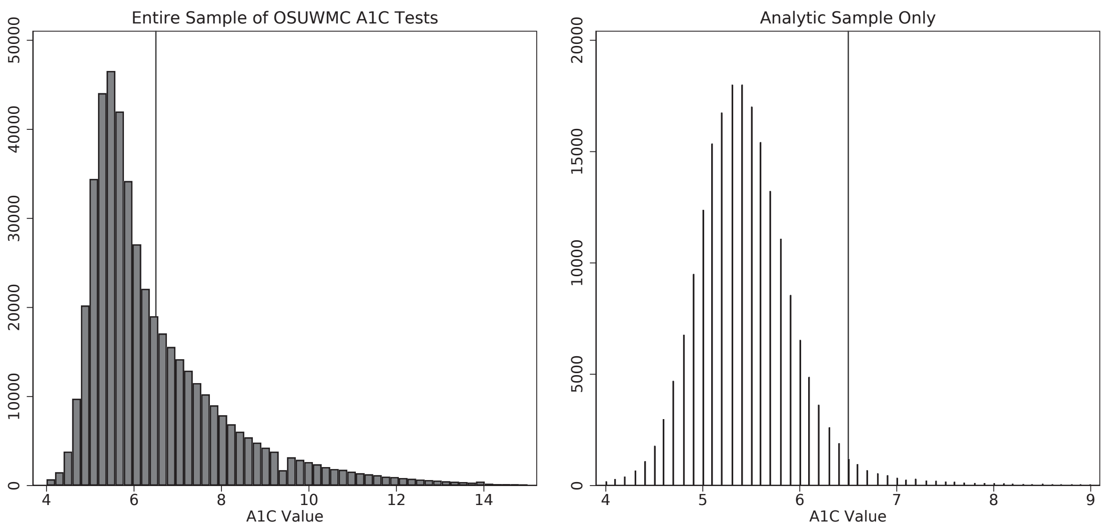

Figure 2.1: Distribution of HbA1C tests at Ohio State Wexner Medical Center

Note: The left panel of this figure presents the frequency distribution of HbA1C tests that a patient received in a given quarter for the entire population of patients receiving tests at the Ohio State University Wexner Medical Center. In the case when a patient received multiple HbA1C tests in the same quarter, the maximum value is plotted. The right panel is limited to the analytic sample of this analysis. In particular, the major restriction is that it removes individuals who have already been diagnosed with diabetes, and hence limits to quarters of those who have never been diagnosed with diabetes or for whom this is the first quarter of treatment. Diagnosis, as described in text, is inferred from the presence of a relevant ICD-10 code or a diabetes-related medication during the quarter. The vertical line is the 6.5% HbA1C threshold where the ADA recommends to begin treatment for diabetes.

	Optum Sample	UCLA Sample	Full OSU Sample	Never Diabetic	Become Diabetic
Female	0.523	0.612	0.577	0.579	0.542
			(0.494)	(0.494)	(0.498)
Age	51.13	47.15	49.486	49.127	55.791
			(15.945)	(15.947)	(14.535)
White	N/A	0.654	0.758	0.761	0.699
			(0.429)	(0.427)	(0.459)
Black	N/A	0.048	0.185	0.182	0.251
			(0.389)	(0.385)	(0.434)
Asian	N/A	0.105	0.045	0.045	0.039
			(0.207)	(0.208)	(0.193)
Ever Medicare	0.281	0.181	0.210	0.202	0.361
			(0.407)	(0.401)	(0.480)
Ever Private	0.719	0.749	0.631	0.638	0.510
			(0.482)	(0.480)	(0.500)
Ever Medicaid	0	0.046	0.159	0.157	0.184
			(0.365)	(0.364)	(0.387)
First Observed Credit Score	N/A	N/A	705.346	706.554	684.145
			(111.260)	(110.998)	(113.715)
N	142541	23882	99723	94348	5375

Table 2.1: Demographic summary statistics of patient population

Note: The three columns to the right of this table compares demographic characteristics of individuals in the OSU patient sample with individuals in the two samples (Optum and UCLA) used in Alalouf et al. (2024). The middle column refers to the full sample that is used in this analysis, and is split out between those who will not become diabetic in this period and those who will. The first two columns are copied from Alalouf et al. (2024) and so are left as N/A where congruent metrics are not available.

A few differences between the treatment and comparison groups are immediately obvious. Most notably, the average age of the comparison group is over six years younger than the diabetic group. There are few differences in the genetic and racial makeup of the two groups, and the rate of being on Medicare is about double that of the control group and about three percentage-points larger for being on Medicaid, indicating that the diabetic population may be older and less affluent, consistent with prior research. A second notable difference between the sample in this analysis and samples used in Alalouf et al. (2024) is the Medicaid population in this study. Because the authors use an insurance claims database, the rate of Medicaid is very low. Conversely, in this analysis, the percentage of patients on Medicaid, about 16% is nearly the rate of Medicaid receipt for the state of Ohio as a whole, making the findings here more generalizable to the population as a whole.[21] Another stark difference is in the mean credit score between the two groups, which is more than 20 percentage-points greater in the comparison group (unadjusted for age). In comparison to Alalouf et al. (2024), the populations are comparable in terms of gender, age, and racial makeup, though the rate of private insurance appears to be greater in both samples of the authors' article than in the Ohio State University sample studied in this essay.

Finally, Table 2.2 shows the distribution of diagnosis frequency over time. We can see that the diagnosis frequency is relatively consistent over time, aside from a decline in Q2 of 2020, which is when medical facilities cancelled or postponed non-emergency medical visits.

[21] https://www.census.gov/library/publications/2023/demo/p60-281.html

	New Diagnoses N	Pct.
Q2 2017	111	1.92
Q3 2017	265	4.58
Q4 2017	310	5.36
Q1 2018	265	4.58
Q2 2018	320	5.53
Q3 2018	287	4.96
Q4 2018	308	5.32
Q1 2019	319	5.51
Q2 2019	304	5.25
Q3 2019	285	4.92
Q4 2019	262	4.53
Q1 2020	338	5.84
Q2 2020	223	3.85
Q3 2020	383	6.62
Q4 2020	393	6.79
Q1 2021	311	5.37
Q2 2021	346	5.98
Q3 2021	399	6.89
Q4 2021	359	6.20
N	5788	

Table 2.2: Diagnosis frequency by period

2.4 Empirical strategy

An ideal identification strategy to consider the economic consequences of diabetes would allow us to observe the consumption, debt repayment, and labor supply of diabetic patients in the counterfactual reality where they had not been diagnosed with diabetes. In reality, diabetes is a challenging disease for clean identification because the onset of the disease might be related to the outcome itself. Thus, an analysis that omits characteristics of the individual that makes disease diagnosis conditionally exogenous to the individual will distort any relationship between disease status and the outcome (i.e. financial health and labor force participation).

2.4.1 Event study difference in differences

Some medical research has documented particular physical characteristics of individuals that are commonly at risk for developing type II diabetes (Wilmot and Idris, 2014; Hillier and Pedula, 2001; Hsia et al., 2009; Haines et al., 2007; Feltbower et al., 2003; Schienkiewitz et al., 2006; Colditz et al., 1995). These characteristics include a high Body Mass Index (BMI), a sedentary lifestyle, being from a socially disadvantage group, and genetic predisposition.[22] However, even given these documented differences between diabetics and non-diabetics, *trends* in variables related to consumption, debt repayment, and labor for patients who will be diagnosed with diabetes are notably similar to trends for patients who have HbA1C labs drawn at the same hospital system but will not become diabetic in the study period. I discuss this with greater specificity below.

[22]Barbieri and Nguyen (2022) exploit genetic pre-disposition as an instrumental variable to study labor supply. The authors find no impact of diabetes on labor supply.

Table 2.3 reports the means and standard deviations of the credit and labor outcomes only for individuals who will be diagnosed with diabetes during the study period, where 0 is the quarter of diagnosis. Examining the first row of Table 2.3, Credit Score, for example, we can see that the four quarters leading up to diagnosis t, those who will be diagnosed with diabetes show a relatively steady increase in credit score of about 1.6-points per quarter between $(t-4)$ and $(t-1)$ (and statistically different at $p < 0.05$). After diagnosis, i.e. between the periods $(t+1)$ and $(t+4)$, diabetic patients show essentially unchanged credit score (and are not statistically different from each other).

As an illustration, consider Figure 2.2, which plots conditional means of credit score over event time for individuals who will be diagnosed in cyclamen, which corresponds to what is presented in Table 2.3. I likewise create event time for individuals who have labs drawn at OSU but will not be diagnosed in our sample period (shown in black), by selecting a random period as event time 0, and defining event time around this, as if they would be diagnosed. Because this is a fake event, the trends of actual diabetic patients and the counterfactuals should move in sync before diagnosis, and any divergence after diagnosis is likely due to the diagnosis itself. First, we can see that the parallel trends assumption seems remarkably satisfied empirically. However, a deeper consideration provides the first empirical evidence that there is perhaps divergence post diagnosis, which is what I test econometrically throughout the rest of the essay.

Figures 2.3-2.6 depict the means over event time in graphical representation. In Figure 2.3, we can see visually the leveling off of credit score growth around diagnosis, as well

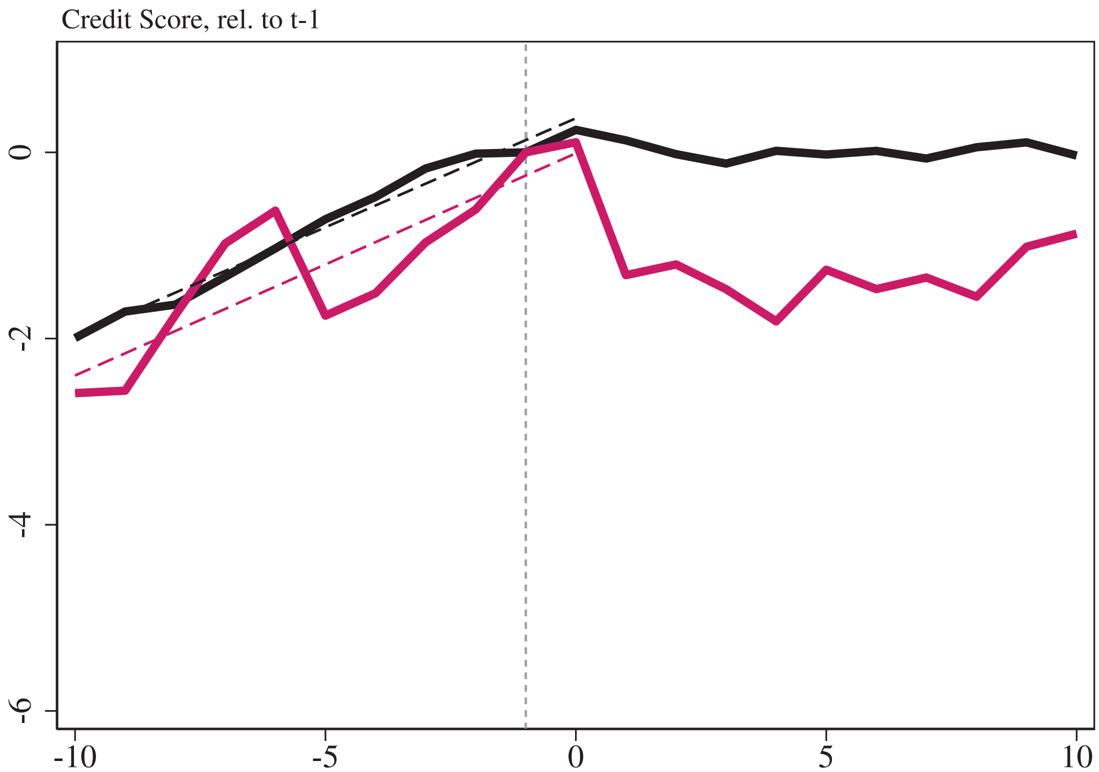

Figure 2.2: Visualizing differences in trends in credit score

Note: This figure plots the change in credit score relative to t-1 in event time for the newly-diagnosed diabetes group (cyclamen), and the same for the never-diabetic group where event time is randomly generated (black). These results are generated from a regression of the form $CreditScore_{i,t} = \alpha + \sum_{k=-19}^{-2} \beta_k event_{i,t} \times 1[Diabetes_i] + \sum_{k=0}^{19} \beta_k event_{i,t} \times 1[Diabetes_i] + age_{i,t} + \theta_i + year_t + quarter_t + \epsilon_{i,t}$. In this regression, $Diabetes$ is an indicator for whether the patient is eventually-diabetic. $Event$ is either the diagnosis quarter for diabetes patients, or the randomly selected diagnosis quarter for non-diabetics. While these results provide visual evidence that trends for eventual-diabetes and non-diabetics are approximately parallel before disease onset, as well as evidence of divergence from the trend of those not affected with diabetes after diagnosis, this visual does not allow for statistical inference. Though it is generated from real data, it is for illustrative purposes of these two characteristics.

as some suggestive evidence around the intensity of credit score change, in particular in the likelihood of having a super prime credit score (720+). In Figure 2.4, the means are less clean for several metrics, but the slowed decline in credit availability, and approximately $700 drop in non-housing debt after three quarters suggest there could be a reduction in consumption. Though small, there is graphical evidence in Figure 2.5 of change in delinquency and charge off behavior around diagnosis (0.8 and 0.7%-points, respectively) and a very small (perhaps negligible, economically) immediate increase in medical collections. Finally, though we see some trends in labor supply, it is difficult to necessarily attribute any obvious changes to diagnosis, as shown in Figure 2.6.

Yet obviously Table 2.3 and Figures 2.3-2.6 tell just part of the story. While the comparison made here provides suggestive evidence that diabetes diagnosis may slow growth in credit access, reduce consumption, and increase default on debt payments, it is necessary to report this relative to the untreated group. Figure 2.2 visually depicts this, and the difference in differences strategy that I employ computes this analytically.

I implement an event study difference in differences of the following form to estimate the impact of diabetes diagnosis on a series of consumption, debt repayment, and labor supply metrics:

$$Y_{i,t} = \alpha + \sum_{k=T_0}^{-2} \beta_k diag_{i,t} + \sum_{k=0}^{T_T} \beta_k diag_{i,t} + \theta_i + \gamma_t + \epsilon_{i,t}$$

The primary specification considers the quarter as the time unit, t. The maximum number of quarters that we can observe an individual is 18 quarters. I include fixed

	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12
	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD
Credit Access:																	
Credit Score	683.786	685.496	687.595	689.876	691.372	691.371	691.302	692.203	691.375	692.429	692.917	693.944	694.838	696.424	698.223	701.141	701.529
	(113.799)	(113.238)	(112.488)	(111.466)	(110.978)	(111.277)	(111.274)	(110.580)	(110.554)	(110.689)	(110.535)	(109.131)	(109.734)	(108.640)	(107.101)	(106.575)	(106.928)
Credit Score < 620	0.331	0.324	0.313	0.301	0.298	0.294	0.292	0.286	0.285	0.278	0.279	0.274	0.271	0.265	0.259	0.246	0.256
	(0.471)	(0.468)	(0.464)	(0.459)	(0.457)	(0.456)	(0.455)	(0.452)	(0.451)	(0.448)	(0.448)	(0.446)	(0.445)	(0.441)	(0.438)	(0.430)	(0.436)
Credit Score 720+	0.429	0.435	0.438	0.443	0.451	0.446	0.447	0.441	0.441	0.439	0.438	0.437	0.442	0.442	0.445	0.452	0.461
	(0.495)	(0.496)	(0.496)	(0.497)	(0.498)	(0.497)	(0.497)	(0.497)	(0.497)	(0.496)	(0.496)	(0.496)	(0.497)	(0.497)	(0.497)	(0.498)	(0.499)
Credit Score < 580	0.243	0.239	0.232	0.220	0.215	0.210	0.207	0.204	0.205	0.199	0.199	0.195	0.194	0.184	0.176	0.168	0.166
	(0.429)	(0.426)	(0.422)	(0.415)	(0.411)	(0.407)	(0.405)	(0.403)	(0.404)	(0.400)	(0.400)	(0.396)	(0.396)	(0.387)	(0.381)	(0.374)	(0.372)
Consumption:																	
Rev. Credit Availability	32.624	32.114	31.702	30.646	30.604	30.808	30.935	30.604	30.675	30.091	30.499	30.364	29.960	29.387	29.460	28.095	27.474
	(36.047)	(36.398)	(35.897)	(35.498)	(33.869)	(34.638)	(34.889)	(34.694)	(35.255)	(34.030)	(33.956)	(33.639)	(33.739)	(33.722)	(34.151)	(32.619)	(32.063)
Total Non-Housing Debt	19926.345	20294.893	20268.724	20285.661	20288.792	19803.608	19751.587	19816.791	19755.737	19682.039	19323.864	19432.329	19576.576	19330.791	19401.853	18614.164	18285.449
	(39401.565)	(40762.219)	(40949.904)	(40579.764)	(39732.292)	(39646.848)	(40137.175)	(39342.314)	(35799.712)	(36665.617)	(34795.455)	(35400.801)	(36631.518)	(35799.007)	(38284.895)	(38449.452)	(34152.421)
Any New Credit Inquiry in Q	0.217	0.218	0.213	0.211	0.218	0.210	0.212	0.212	0.208	0.212	0.197	0.199	0.196	0.202	0.189	0.178	0.186
	(0.412)	(0.413)	(0.410)	(0.408)	(0.413)	(0.407)	(0.408)	(0.409)	(0.406)	(0.409)	(0.398)	(0.399)	(0.397)	(0.402)	(0.391)	(0.383)	(0.390)
Any New CC Trade in Q	0.050	0.059	0.063	0.061	0.066	0.062	0.059	0.062	0.061	0.053	0.055	0.056	0.054	0.055	0.049	0.052	0.050
	(0.218)	(0.236)	(0.243)	(0.240)	(0.248)	(0.241)	(0.236)	(0.241)	(0.239)	(0.225)	(0.228)	(0.229)	(0.227)	(0.229)	(0.216)	(0.223)	(0.218)
Any New Auto Trade in Q	0.033	0.032	0.031	0.034	0.032	0.033	0.031	0.037	0.030	0.030	0.029	0.030	0.031	0.037	0.028	0.027	0.028
	(0.178)	(0.176)	(0.172)	(0.182)	(0.176)	(0.178)	(0.173)	(0.189)	(0.169)	(0.172)	(0.168)	(0.172)	(0.173)	(0.188)	(0.165)	(0.163)	(0.164)
Any New Mort Trade in Q	0.012	0.015	0.013	0.015	0.012	0.013	0.013	0.012	0.015	0.014	0.013	0.011	0.013	0.012	0.013	0.019	0.012
	(0.111)	(0.120)	(0.115)	(0.123)	(0.111)	(0.115)	(0.113)	(0.107)	(0.120)	(0.119)	(0.114)	(0.104)	(0.114)	(0.107)	(0.114)	(0.138)	(0.107)
Repayment History:																	
Any 60+ Day Delinquency in Prev Y	0.138	0.136	0.130	0.124	0.127	0.132	0.132	0.133	0.134	0.133	0.133	0.131	0.132	0.130	0.119	0.106	0.106
	(0.345)	(0.342)	(0.336)	(0.330)	(0.333)	(0.339)	(0.339)	(0.339)	(0.341)	(0.340)	(0.339)	(0.337)	(0.338)	(0.337)	(0.324)	(0.308)	(0.307)
Any Discharged BK in Q	0.001	0.002	0.003	0.001	0.003	0.000	0.001	0.001	0.002	0.002	0.001	0.001	0.002	0.002	0.003	0.002	0.002
	(0.035)	(0.048)	(0.054)	(0.035)	(0.053)	(0.020)	(0.038)	(0.033)	(0.041)	(0.039)	(0.034)	(0.025)	(0.049)	(0.043)	(0.054)	(0.048)	(0.041)
Any Charge off in Q	0.085	0.085	0.081	0.078	0.079	0.080	0.081	0.081	0.082	0.084	0.084	0.082	0.081	0.080	0.074	0.069	0.066
	(0.279)	(0.279)	(0.273)	(0.268)	(0.269)	(0.271)	(0.272)	(0.273)	(0.274)	(0.277)	(0.278)	(0.274)	(0.273)	(0.271)	(0.262)	(0.253)	(0.249)
Any Foreclosure in Q	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000
	(0.016)	(0.026)	(0.025)	(0.025)	(0.024)	(0.020)	(0.014)	(0.021)	(0.022)	(0.023)	(0.024)	(0.030)	(0.032)	(0.019)	(0.000)	(0.022)	(0.000)
Total Medical Collections	322.796	318.445	316.101	314.664	290.281	299.473	288.246	273.724	269.387	256.524	227.785	231.666	238.876	224.773	215.571	198.100	208.086
	(1713.254)	(1645.465)	(1582.506)	(1589.384)	(1535.391)	(2080.254)	(2006.240)	(1428.652)	(1376.237)	(1444.650)	(1226.572)	(1282.038)	(1629.006)	(1531.799)	(1433.161)	(1401.565)	(1517.149)
CC Availability (0,100)	0.516	0.515	0.526	0.530	0.525	0.523	0.528	0.521	0.525	0.515	0.524	0.519	0.516	0.514	0.516	0.509	0.506
	(0.500)	(0.500)	(0.499)	(0.499)	(0.499)	(0.500)	(0.499)	(0.500)	(0.499)	(0.500)	(0.499)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Labor Supply:																	
LFP	0.470	0.473	0.461	0.451	0.438	0.415	0.420	0.414	0.416	0.396	0.390	0.393	0.382	0.381	0.352	0.335	0.331
	(0.499)	(0.499)	(0.499)	(0.498)	(0.496)	(0.493)	(0.494)	(0.493)	(0.493)	(0.489)	(0.488)	(0.489)	(0.486)	(0.486)	(0.478)	(0.472)	(0.471)
Total Wage Earnings in Q of Worked	13543.707	13673.932	13628.013	14120.901	13712.384	14119.051	14015.410	14733.242	15061.094	14367.697	14711.754	14736.668	14587.822	13926.080	13027.057	12690.911	14655.295
	(14729.765)	(17073.625)	(14436.651)	(19577.392)	(14869.958)	(18077.724)	(15118.128)	(19938.766)	(26773.930)	(15100.617)	(15944.176)	(17320.693)	(16065.116)	(12725.510)	(9082.158)	(8558.069)	(15287.879)
Max Number of Weeks Worked in Q of Worked	11.360	11.535	11.317	11.416	11.350	11.451	11.490	11.429	11.551	11.474	11.478	11.380	11.814	11.590	11.438	11.479	11.992
	(2.822)	(2.628)	(2.989)	(2.890)	(2.829)	(2.810)	(2.782)	(2.759)	(2.662)	(2.768)	(2.938)	(2.943)	(2.465)	(2.512)	(2.769)	(2.772)	(1.896)
Number of Employers in Q of Worked	1.157	1.165	1.144	1.137	1.133	1.152	1.163	1.145	1.153	1.127	1.135	1.185	1.131	1.146	1.119	1.110	1.131
	(0.455)	(0.486)	(0.434)	(0.432)	(0.378)	(0.451)	(0.445)	(0.400)	(0.441)	(0.385)	(0.371)	(0.519)	(0.396)	(0.392)	(0.444)	(0.374)	(0.385)

Table 2.3: Means (SD) of select dependent variables for treated group

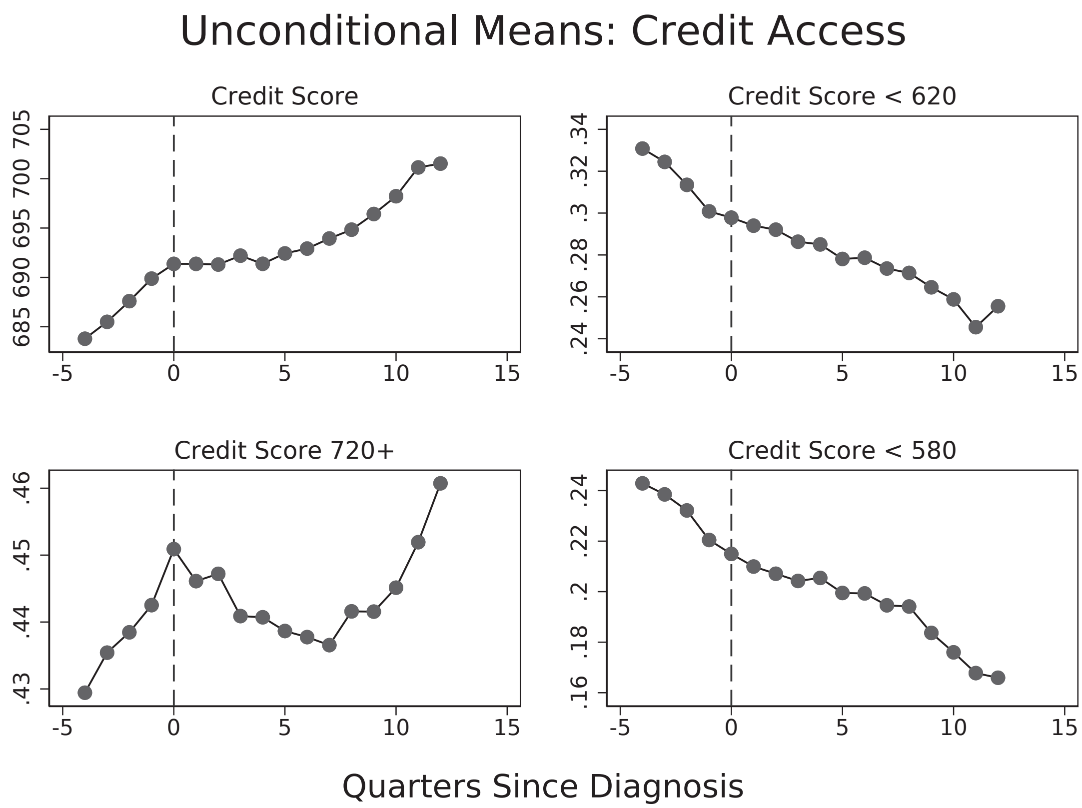

Figure 2.3: Unconditional means of diagnosed group - credit access

Note: The figures presents unconditional means for four variables related to credit access. The top left panel plots credit score. The upper right, bottom left, and bottom right plot the likelihood of having a credit score that is below prime (below 620), super prime (above 720), and deep prime (below 580). These classifications are meaningful classes of credit scores that are used in part to determine the cost of access to credit. The means are limited to individuals who will be newly diagnosed with diabetes during the study period, and are plotted in event time.

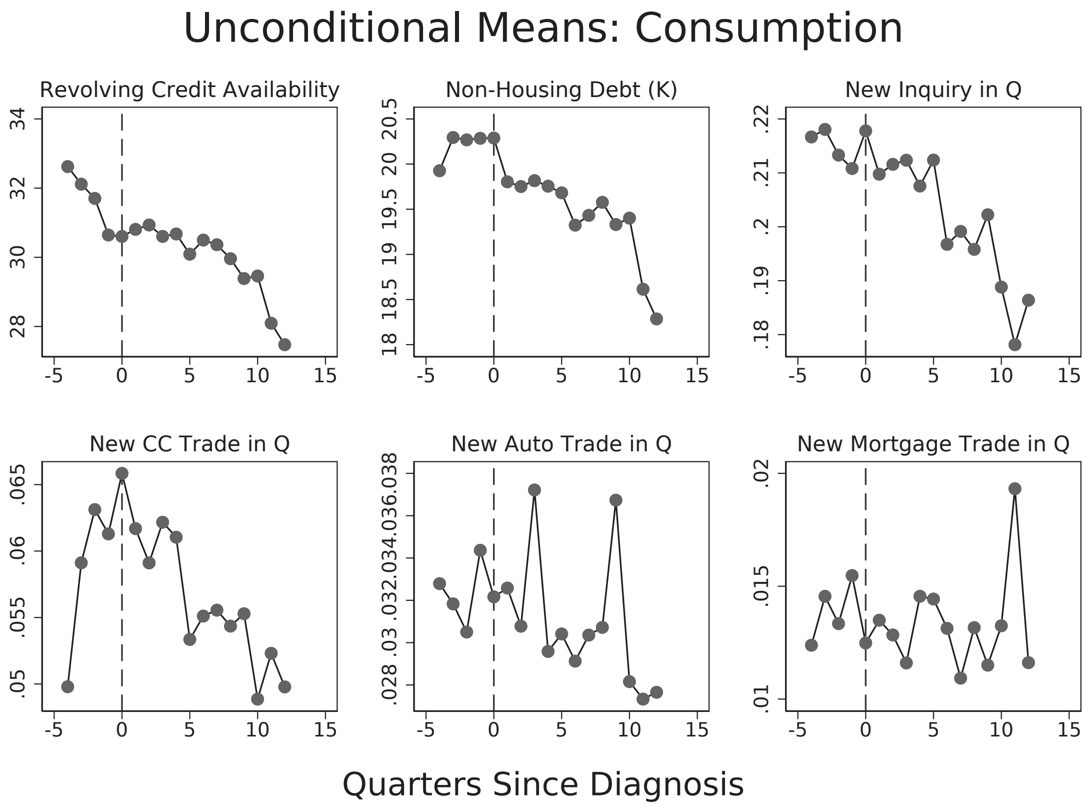

Figure 2.4: Unconditional means of diagnosed group - consumption

Note: The figures presents unconditional means for six variables related to consumption. While consumption is notably challenging to observe in consumer credit data, this figure plots trends in revolving credit availability, non-housing debt, new inquiries in a quarter, and new credit card, auto, and mortgage trades in a quarter. The means are limited to individuals who will be newly diagnosed with diabetes during the study period, and are plotted in event time.

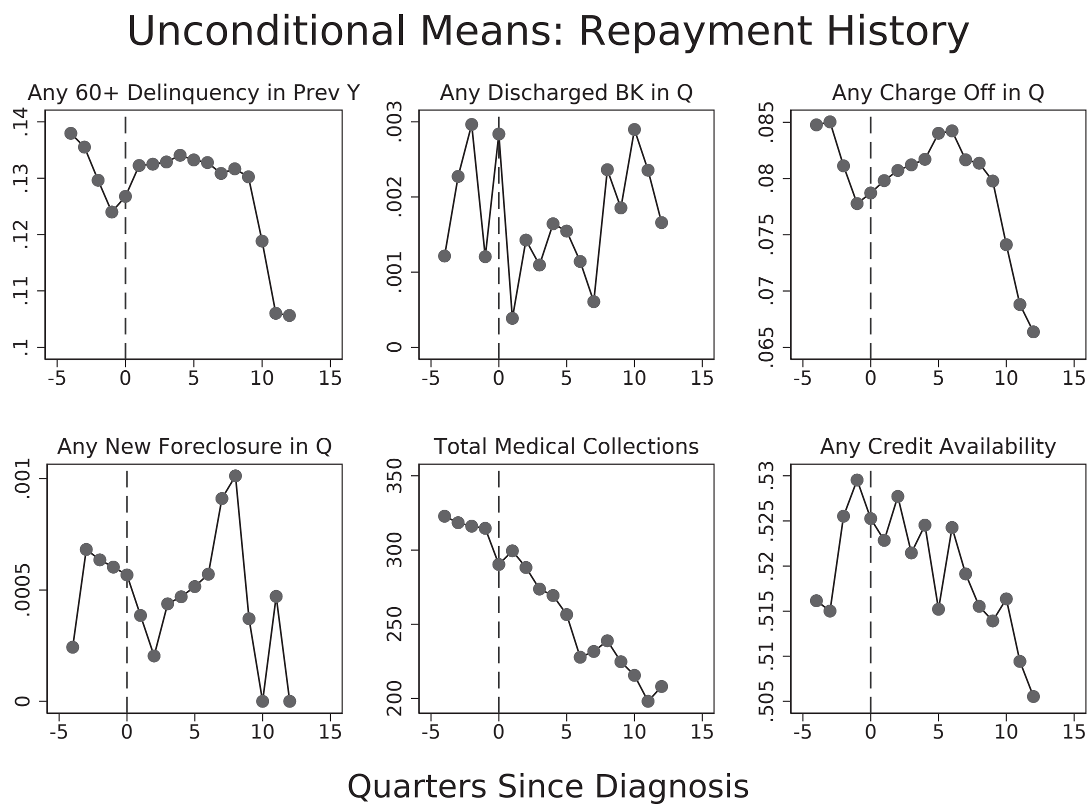

Figure 2.5: Unconditional means of diagnosed group - repayment history

Note: The figures presents unconditional means for six variables related to repayment history. While the exact formula to determine credit score is proprietary, variables related to repayment history are the dominating factor in adverse credit score changes. 60-day delinquencies, discharged bankruptcies, charge offs, foreclosures, medical collections, and likelihood of having any remaining credit availability are plotted. The means are limited to individuals who will be newly diagnosed with diabetes during the study period, and are plotted in event time.

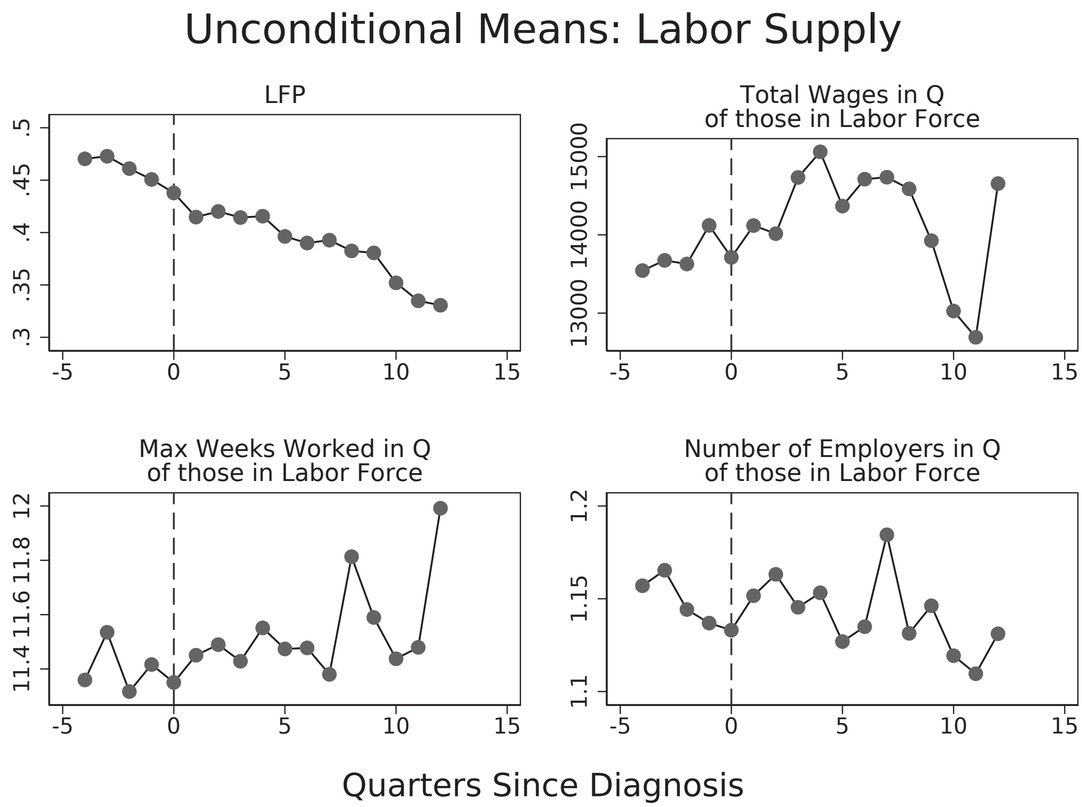

Figure 2.6: Unconditional means of diagnosed group - labor supply

Note: The figures presents unconditional means for four variables related to labor supply. Plotted are the labor participation rate, total wages, weeks of labor supplied, and number of employers in a quarter. The means are limited to individuals who will be newly diagnosed with diabetes during the study period, and are plotted in event time.

effects, θ, for the individual and, γ, for calendar time in an event study difference in differences with event time dummies captured in β_k, where the parallel trends assumption suggests that $\beta_{k<0}$ is statistically zero, and the treatment effect relative to pre-treatment is captured in $\beta_{k\geq 0}$.

2.4.2 Fuzzy regression discontinuity design

While the event study difference in differences provides a picture of the progression of financial outcomes for all individuals diagnosed with diabetes, which allows for an estimate of the ATE, another approach is to isolate differences for those marginally diagnosed relative to those marginally not diagnosed, or an estimate of the LATE. To do this, I use fuzzy RDD in a congruent way to what has been used in previous literature (Alalouf et al., 2024; Gaggero et al., 2022; Iizuka et al., 2021; Kim et al., 2019) with HbA1C as the running variable.

While the estimates from this empirical design should be unbiased estimates of the effect of diabetes diagnosis, the style of model is not without its limitations. The biggest limitation is certainly that the scope of inference provided by this model is limited to the individuals right at the margin, and not more general in scope. Thus, we are unable to generalize the results for patients with more severe diabetes at the time of diagnosis who have HbA1C values beyond the marginally diagnosed threshold. Additionally, as I will show below, while there is a clear discontinuity in the relationship between HbA1C value and probability of diagnosis around the 6.5% threshold recommended by the ADA, the relationship between the running variable (i.e. the distance between the HbA1C value and 6.5%) and the outcome variable (e.g. credit score after one year) is often quite sensitive to modeling choices. I expand on this below, and in Appendix C. Conversely, a null finding in an RDD specification may conceal a real relationship more broadly because it only considers a subsection of the population, in this case those who are the least sick. With the caveat taken into consideration, the major benefits to this strategy in this essay include being able to rigorously separate the effect of the disease vs the diagnosis, and being able to expand almost directly on prior findings that use a similar approach. In short, while there are benefits to using the RDD in assessing the relationship between marginal diabetes diagnosis and financial consequences, it is not a panacea and requires careful – and honest – consideration.

The general strategy that I employ is to exploit the randomness of diagnosis after conditioning on the HbA1C value itself. While 6.5% is the threshold for diagnosis recommended by the ADA, our data, which mirrors the scenario reported in (Alalouf et al., 2024), show that there is not strict compliance by the doctors. Indeed, Figure 2.18 below, also shows this phenomenon clearly. While there is a clear jump at 6.5%, the probability of diagnosis does not increase from 0% below 6.5% HbA1C to 100% above 6.5% HbA1C, which would be the case if there was perfect compliance with the ADA recommendation. Instead, the probability of diagnosis jumps by about 20-percentage points. This could be due to imperfect compliance to the recommendations from the physicians, errors in coding individuals as diabetic in the medical records, or failure to seek follow-up care from the patient (either entirely or at facilities associated with Ohio State University, where we obtain the medical records).[23]

[23] While each of these are possibilities, and probably occur, reimbursement incentives to code patients as diabetic diminish the *a priori* likelihood of the first two possibilities, and possibility three is diminished by insurance networks. In particular, many individuals who receive tests at Ohio State University are insured by the university, and are incentivized to receive care in-network (i.e. at University medical facilities).

Because there is not strict compliance of the 6.5% threshold, this is a scenario for a fuzzy RDD. This can be thought of conceptually as two-stage least squares (2SLS). In practice, the specifications use the RD Robust package, which is the dominant empirically-driven inference method (Calonico et al., 2014).

1S:

$$Diag_{t+n,i} = \beta_1 HbA1C_{t,i} + \beta_2 \mathbb{1} \cdot (HbA1C > 6.5\%)_{t,i} + \beta_3 HbA1C \times \mathbb{1} \cdot (HbA1C > 6.5\%)_{t,i} + \epsilon_{t,i}$$

2S:

$$Y_{t+n,i} = \gamma_1 D\hat{i}ag_{t+n,i} + \upsilon_{t,i}$$

Y is the outcome of study, n is the number of quarters after the test that are considered in the specific model $\{4, 8, 12\}$. $Diag_{t+4,i}$ is simply an indicator of whether the individual was diagnosed within four quarters after the test in period t.

Conditioning on the HbA1C value of the individual allows the instruments β_2 and β_3 to exogenously affect the diagnosis status. As a result, the coefficient of interest in stage two, γ_1, allows us to understand differences in conditional means between the two groups. In particular, the null hypothesis is that the difference between groups is zero, or specific to these visualizations, that the error bar for the difference in means does not intersect with 0. The corresponding tables to the visualizations are available upon request from the author.

Finally, this design implicitly controls for the lifestyle differences that are associated with diabetes. It exploits the imperfect compliance of diabetes diagnosis by conditioning on the HbA1C test value. Hence, the interval validity of these estimates is fairly strong for patients right on the margin of being diagnosed; however, since the estimates are generated by marginal changes in HbA1C values, the generalizability to all diabetes patients is limited.

2.5 Main results

2.5.1 Event study difference in differences

Consumption

The first group of dependent variables that I consider are those related to consumption, shown in Figure 2.7. Generally, I find little evidence that consumption differs after diagnosis for diabetic patients, relative to non-diabetic patients. This finding is supported by Chetty and Szeidl (2007) who show that a household's ability to moderate its spending month to month is inflexible due to high degree to which modern consumption is regularly recurring. Hence a household may be unable to reduce its consumption. There is, however, pronounced decrease in credit utilization, which suggests that individuals may be consuming less, assuming their level of credit has remained the same. I also find some evidence of a reduction in non-housing debt, but the proximity from diagnosis (3+ years) gives rise to skepticism of its relation to diabetes diagnosis alone. The only other metric that appears to offer some evidence for change in consumption behavior is relatively consistent declines in originating a new mortgage trade within the first two years of diagnosis. While this is a question with plausible supply-side and demand-side explanations, I find generally that patients are approximately 40% less likely to originate a new mortgage trade after eight quarters relative to the baseline.

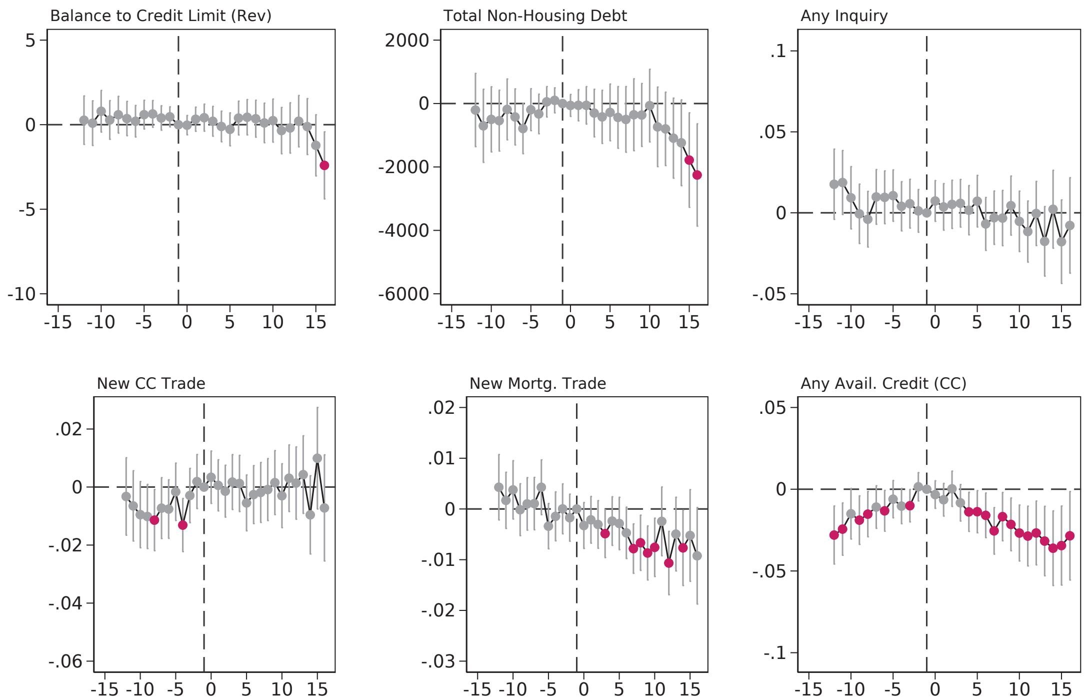

Figure 2.7: Difference in differences results - consumption

Note: This figure visualizes the results from the event study difference in differences model. Plotted on the x-axis are quarters since diagnosis, centered on zero. Coefficient estimates and 95% confidence intervals are displayed. When the coefficient is statistically different from zero (at α=0.05), the coefficient is plotted in cyclamen. Results are displayed for three years prior to diagnosis and four years post diagnosis, though models are estimated for 18 quarters on either direction. Full results are found in Appendix A.

Debt repayment

Because there appears to be little change in consumption aside from decreased credit utilization, next I explore the possibility of changes in debt repayment. Figure 2.8 delves into the debt repayment outcomes. I find a small, though significant, sustained increase in having any 60+ delinquencies reported in the last year and a decline in likelihood of having revolving credit in the first two years after diagnosis. One caveat to this finding is the existence of small pre-treatment trends in delinquency and credit availability about two years before diagnosis. These pre-trends are actually in the opposite direction of the effect, which suggests that the imperfect counterfactual in this scenario would actually be understating the magnitude of the effect. Still, it is worth noting this caveat in the results. While in aggregate there is no statistical evidence of an increase in chargeoffs in the quarters after diagnosis, as I will note when discussing heterogeneity in the results below, there is a staggering five percentage point increase (or roughly 100% of the baseline chargeoff rate) in the presence of chargeoffs immediately following diagnosis for individuals who are credit constrained.

Credit Performance

Recall that the two single most important factors in the FICO credit score are payment history and credit utilization, and so the findings shown in Figures 2.7 and 2.8 offer support for the decline in credit score discussed above. Furthermore, because there were only small changes in repayment history and credit utilization, we might expect that changes to credit score will be small.

The results in Figure 2.9 are labeled as "Credit Performance", which treats credit score as an aggregate measure of credit repayment; however, credit score can also be

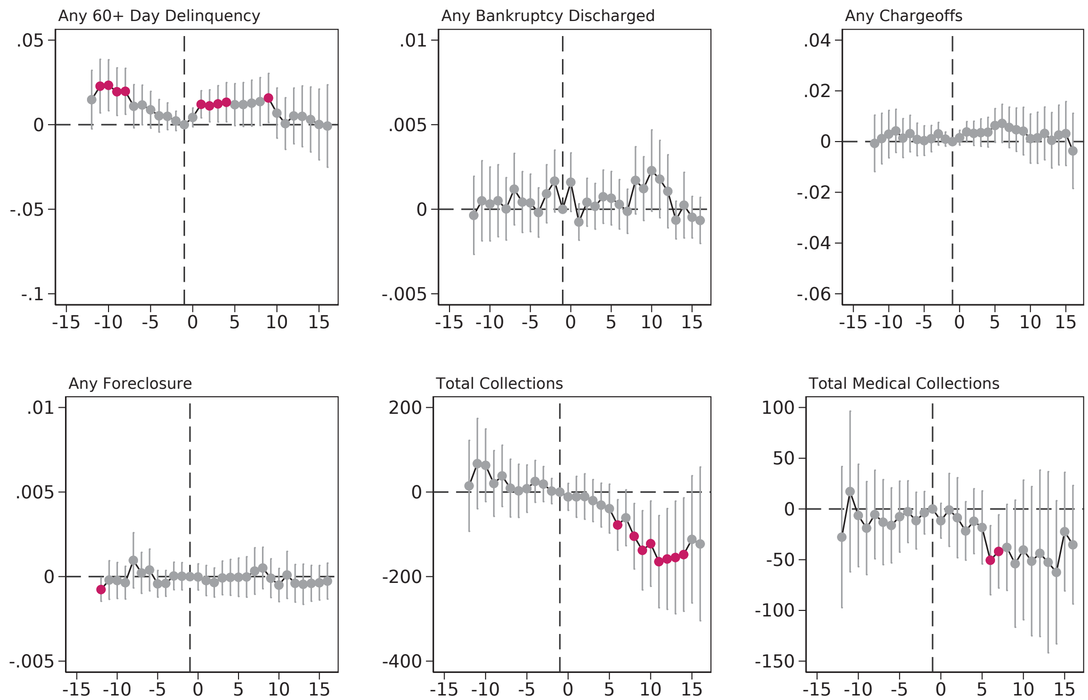

Figure 2.8: Difference in differences results - repayment history

Note: This figure visualizes the results from the event study difference in differences model. Plotted on the x-axis are quarters since diagnosis, centered on zero. Coefficient estimates and 95% confidence intervals are displayed. When the coefficient is statistically different from zero (at α=0.05), the coefficient is plotted in cyclamen. Results are displayed for three years prior to diagnosis and four years post diagnosis, though models are estimated for 18 quarters on either direction. Full results are found in Appendix A.

considered as a major component in the cost of obtaining additional credit in the future, and so a comparable term sometimes employed by scholars is "Credit Access." Accordingly, I analyze credit score by itself in the upper left panel of Figure 2.9, and I also try to dive a bit deeper into the implications of the observed dropped in credit score in the remaining panels of Figure 2.9.

We can see a small, though precisely estimated decline in credit score immediately following diagnosis, which appears to be sustained for the majority of the subsequent quarters. Even where the decline in credit score is insignificant at quarters nine through eleven post-diagnosis, the point estimates are consistent with a sustained decline. Though I find clear evidence that credit does decline, the economic meaning of the decline is less stark. On average, I observe a decline of about 1.75 points in the eight quarters after diagnosis. Moreover, I find no evidence that diagnosis is pushing individuals into subprime (< 620) or deep subprime (< 580) credit buckets. However, I do find that the likelihood of having a superprime (720+) credit score immediately and consistently declines in the quarters following diagnosis. After three years, the likelihood of having a super prime credit score has fallen by approximately three percentage points, or about 6% relative to the baseline probability.

I also analyze potential changes to the likelihood that credit score lies within the subprime (580-619) and prime (660-719) credit score classifications. I find no change in the likelihood of having a subprime credit score. There is a very small immediate decline of about half a percentage point in the likelihood of having a prime credit score, but this effect is not sustained and is generally statistically insignificant and near zero. Results for these models are in the Appendix B. Taken together, the

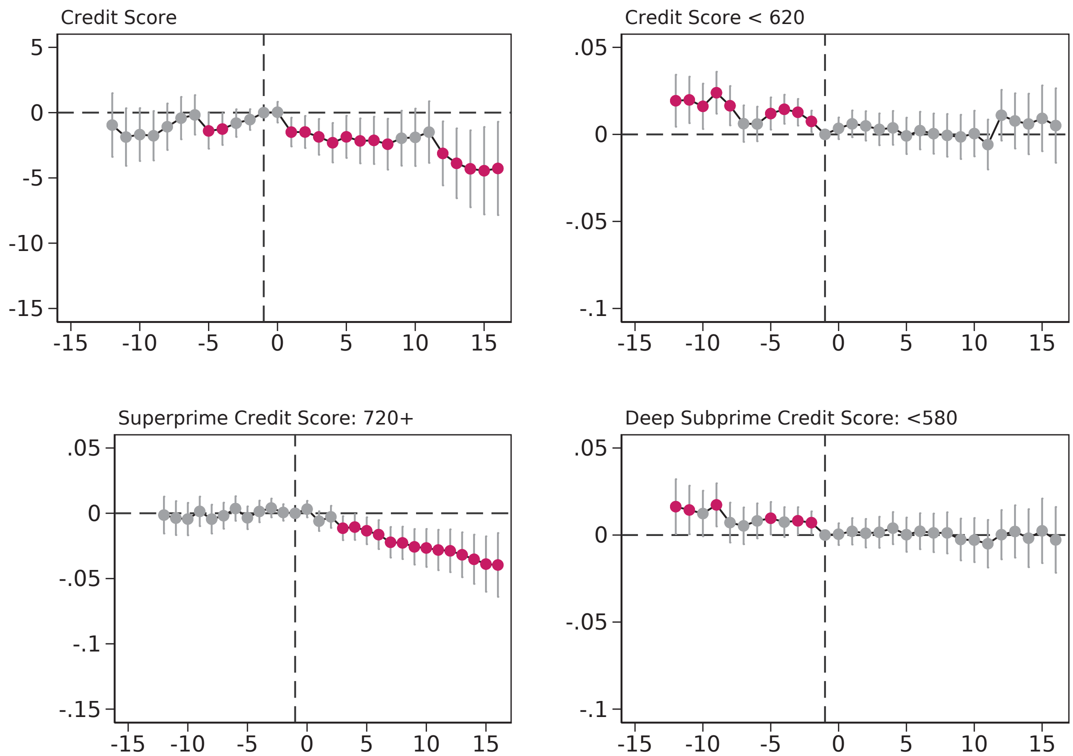

Figure 2.9: Difference in differences results - credit performance

Note: This figure visualizes the results from the event study difference in differences model. Plotted on the x-axis are quarters since diagnosis, centered on zero. Coefficient estimates and 95% confidence intervals are displayed. When the coefficient is statistically different from zero (at α=0.05), the coefficient is plotted in cyclamen. Results are displayed for three years prior to diagnosis and four years post diagnosis, though models are estimated for 18 quarters on either direction. Full results are found in Appendix A.

results here suggest that aggregate drops in credit scores are not being driven by individuals falling into deep subprime and subprime credit classifications, but rather that diabetes diagnosis is increasing the cost of credit for individuals who previously had the cheapest access to credit.

Labor supply

Finally, I consider changes in labor supply, shown in Figure 2.10. As previously noted, the labor response in the context of medical disease is theoretically ambiguous. Figure 2.10 suggests an unambiguous (and sustained) decline in labor force participation almost immediately following diagnosis. At the trough, the likelihood of being employed after three-years is approximately 7% less than pre-diagnosis, comprising approximately a 13% decline in labor supply relative to the baseline. The visualization in Figure 2.10 does call attention to the characteristic that labor supply, even pre-diagnosis, does seem to be on a negative trajectory. The consideration of statistical significance can sometimes obscure the bigger picture that event study can offer.[24] The remaining panels of Figure 2.10 report the unconditional event time estimates in the maximum number of weeks worked in a quarter, the total wage earnings in a quarter, and the number of employers in a quarter. I find declines in all of the three outcomes, consistent with the findings of labor force participation, though I note potential pre-trends 2-years from diagnosis particularly in wage earnings.

Taken as a whole, the findings in this section provide suggest a more pronounced decline in debt repayment than in consumption when diagnosed with type II diabetes, though I note a decline in credit utilization. When I assess the economic impact of this increased likelihood to be delinquent on debt, I find only small impacts to credit

[24] Rambachan and Roth (2023) make the point that the crutch that scholars often lean on with statistical significance is that when confidence intervals overlap with zero, the true effect could reasonably fall anywhere within the band. Accordingly, authors looking for publishable results are generous in removing pre-trends by arguing that coefficient estimates are as close to zero as possible, but stingy in recognizing that this same argument could be used to exacerbate the pre-trend. In doing so, the broad trend is sometimes concealed. Hence, while the visual of LFP in Figure 2.10 reveals a stark decrease after diagnosis, a specific consideration of labor supply in itself should integrate the tools outlined in Rambachan and Roth (2023) for sensitivity to inference on pre-trend assumptions. This also suggests that a secondary empirical strategy could be useful to add robustness to the results.

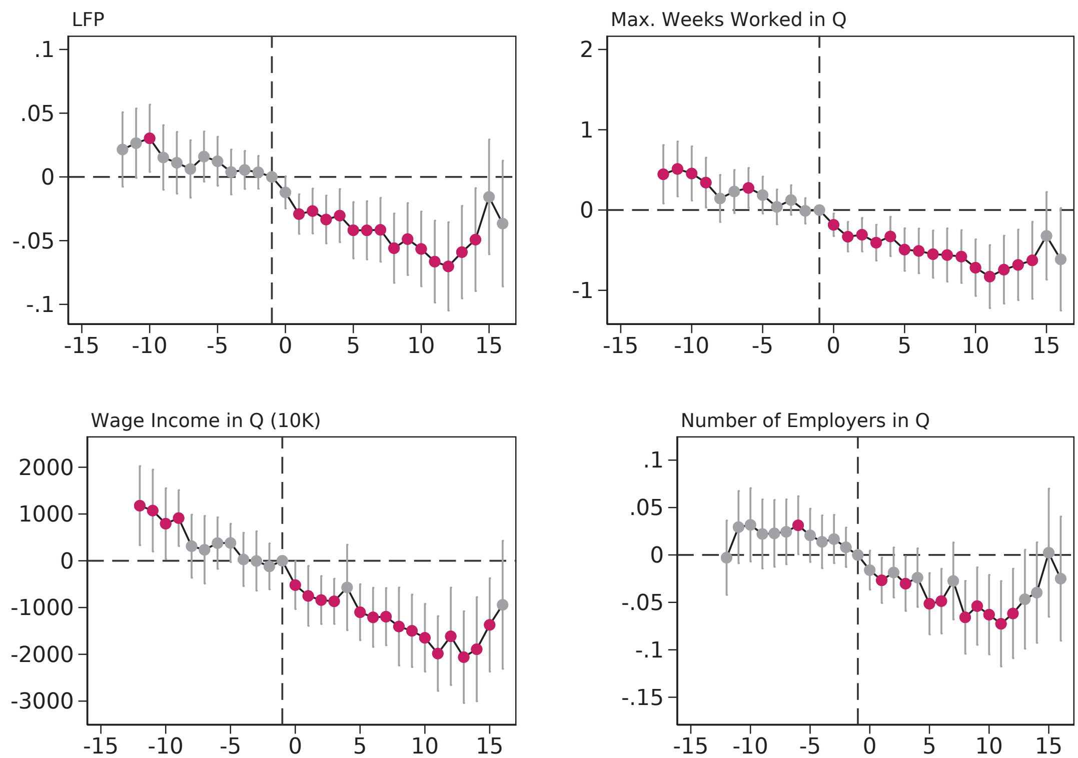

Figure 2.10: Difference in differences results - labor supply

Note: This figure visualizes the results from the event study difference in differences model. Plotted on the x-axis are quarters since diagnosis, centered on zero. Coefficient estimates and 95% confidence intervals are displayed. When the coefficient is statistically different from zero (at α=0.05), the coefficient is plotted in cyclamen. Results are displayed for three years prior to diagnosis and four years post diagnosis, though models are estimated for 18 quarters on either direction. Full results are found in Appendix A.

score. In particular, I estimates a sustained 2-4 point decline in credit score, driven by increases in delinquency, charge offs, and credit utilization. This change in credit score does not appear to be driving individuals to sub prime credit score categories, which suggests that declines in credit *scores* are not necessarily resulting in decreased *access* to credit. Instead, I find that diabetes diagnosis may reduce the likelihood of having a premium credit score, and hence increase the cost of accessing credit for those with the lowest cost of credit. In particular, I find that individuals are 6% less likely to have a super prime credit score relative to baseline. This impact appears to be sustained and does not curtail over time.

The results on labor supply seem to support the idea that chronic illness may be very time costly, since the results that I find show an immediate and sustained decreased in labor supply. While there is rational justification for this decline in labor supply, the results seem to only exacerbate the interpretation of financial distress. An individual who leaves the labor force while incurring medical costs seems to be in a worse financial state than one who is able to positively adjust his labor supply in response, which appears to be the case here.

2.5.2 Heterogeneity in event study results

Finally, I turn to heterogeneity in the main difference in differences results that I presented above. I explore heterogeneity in age, in household type, and in credit availability at diagnosis. I provide commentary on select results, categorized by construct, and provide the full set of results in Appendix B to this chapter.

Age

Figure 2.11 considers rate of delinquency by age, and shows that the effect is concentrated in older adults. We see a five percentage point increase in the rate of having a 60 day or more delinquency in the last year. We can also see a substantial decrease in credit availability and corresponding decrease in balance to credit limit in 2.12. As noted, both of these factors are significant components in the credit score calculation, and thus validate the change in credit score classification. Accordingly, Figure 2.13 delves into heterogeneity in the main result of a change in the relative cost of credit. We can see that, while there is no change in the likelihood of having a superprime credit score after diagnosis for the young group, for the older group, there is a immediate and substantial decline that does not appear to taper. We similarly see an increase in the likelihood of an increase of having a below prime and deep subprime credit scores. Though not shown, I do not find notable labor effects for older nor younger patients.

Household type

In comparing patients who are single at diagnosis to patients who are not single at diagnosis, we can see that the decline in labor supply is concentrated in the non-single individuals. Those who are single appear to remain in the labor force, working approximately the same number of weeks and earning statistically similar amounts of wages per quarter. This is an intriguing result, because it suggests that the household may indeed act as a buffer against the challenges of chronic illness. A patient who is single may be unable to adjust his labor, owing to added financial strain associated with the disease, whereas those who have household members may have the luxury of

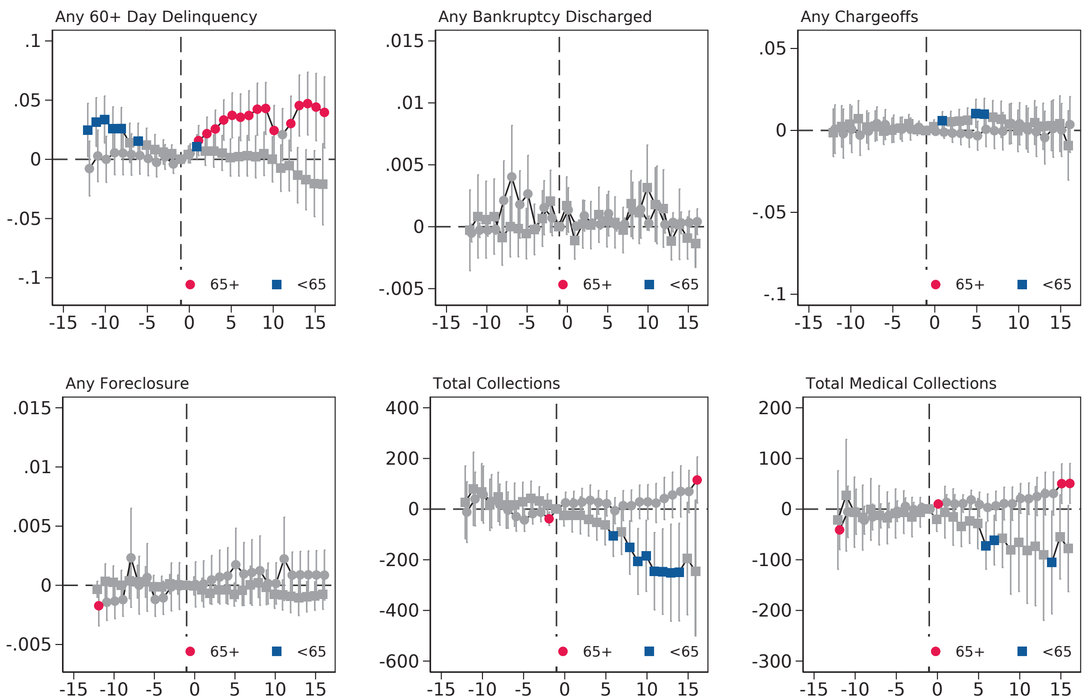

Figure 2.11: Repayment history results by age group

*Note:*The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

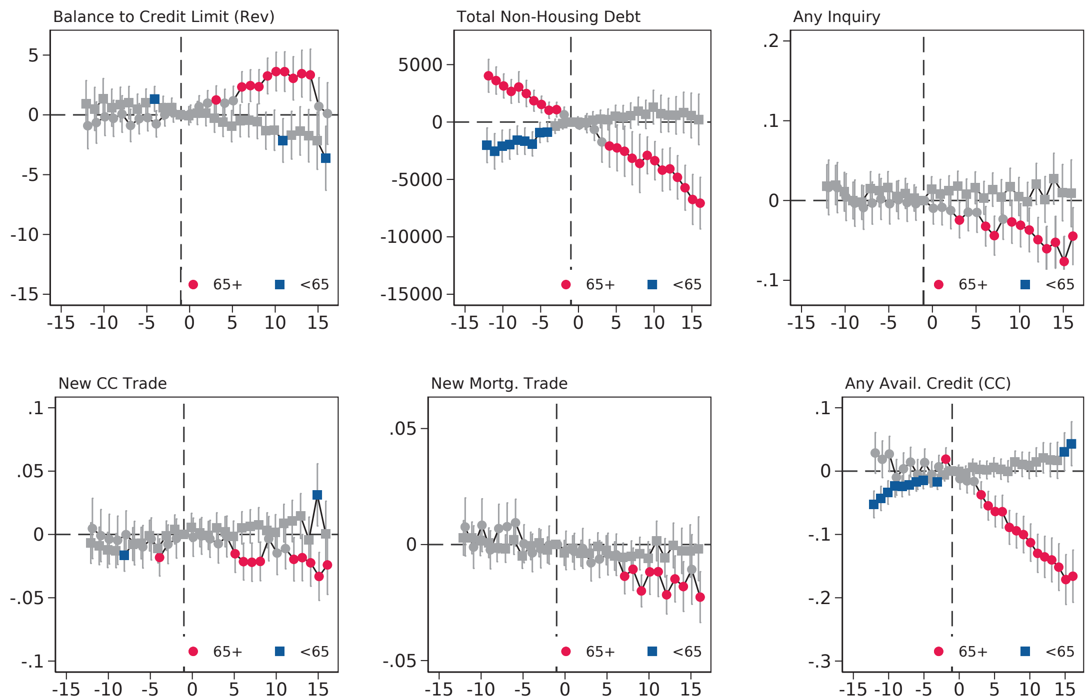

Figure 2.12: Consumption results by age group

*Note:*The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

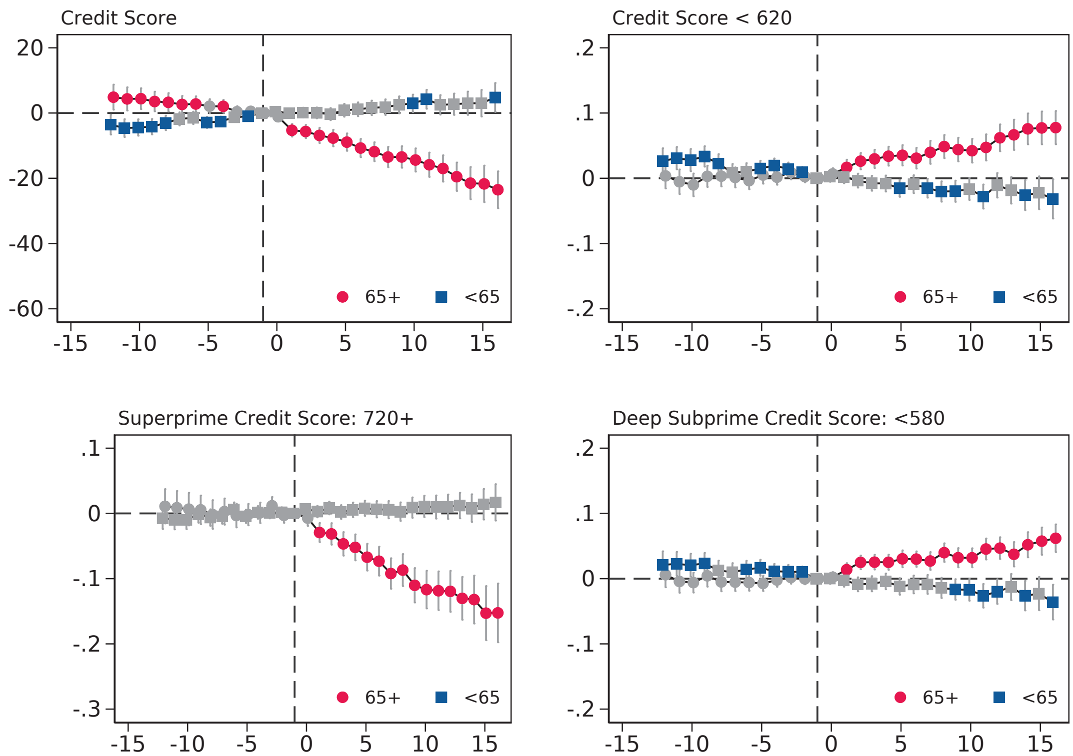

Figure 2.13: Credit performance results by age group

Note: The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

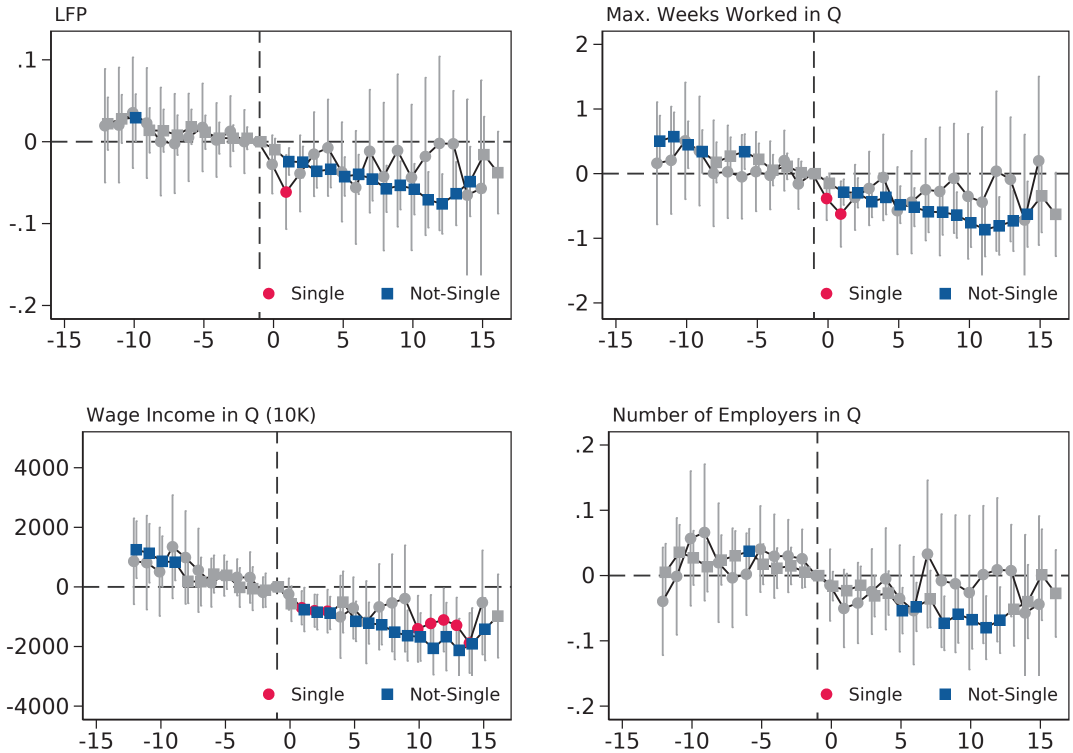

Figure 2.14: Labor supply results by household type

Note: The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

being able to adjust their labor supply. One potential caveat to this is that individuals who are single may be more likely to be older, and hence already supply reduced amount of income. Therefore, an alternative explanation might be that the lack of change relative to pre-diagnosis might really just signal a change from none to none. Though not shown, I do not find notable differences in consumption nor debt repayment by household type.

Credit availability

Finally, turning to heterogeneity in the main repayment history results by the group of patients who are credit constrained vs those who are not, we can see an increase in the delinquency rate for both groups. This appears to be sustained in a more pronounced way for those with a high credit utilization ratio. Most striking, though is the increase in chargeoffs that is concentrated in the group of patients who are credit constrained. The increase of about five percentage points relative to pre-diagnosis occurs in the first quarters after diagnosis and remains stable for at least three years. Consumption results in Figure 2.16 are difficult to access because the definition of heterogeneity is related to several of the consumption outcomes, however, there are no stark differences in consumption results that would suggest a consumption response. While we see an increase in delinquency in both credit constrained and non-credit constrained patients, and pronounced increases in charge offs for the credit constrained, the results in Figure 2.17 suggest that the effects on credit performance and future cost of credit access are concentrated in the non-credit constrained. We see a large decrease in credit score after three years, which is comprised of a decrease in the likelihood of having a superprime credit score, and increases in having below prime and deep subprime credit score. Though not shown, I do not find notable differences in labor supply by credit availability.

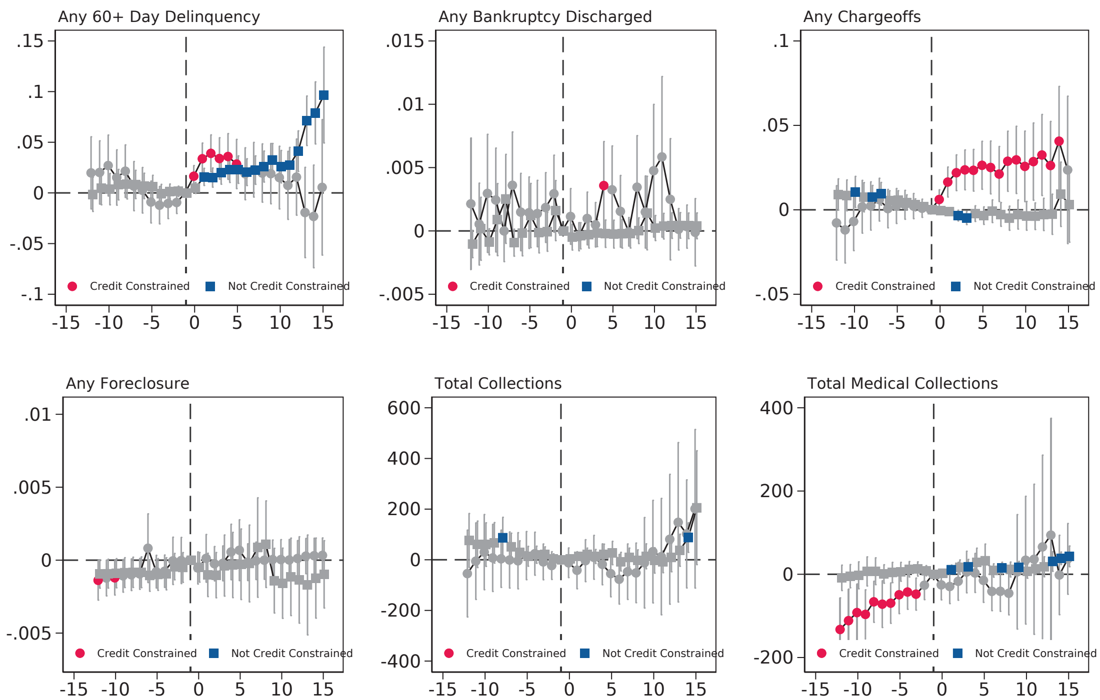

Figure 2.15: Repayment history results by credit availability

*Note:*The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

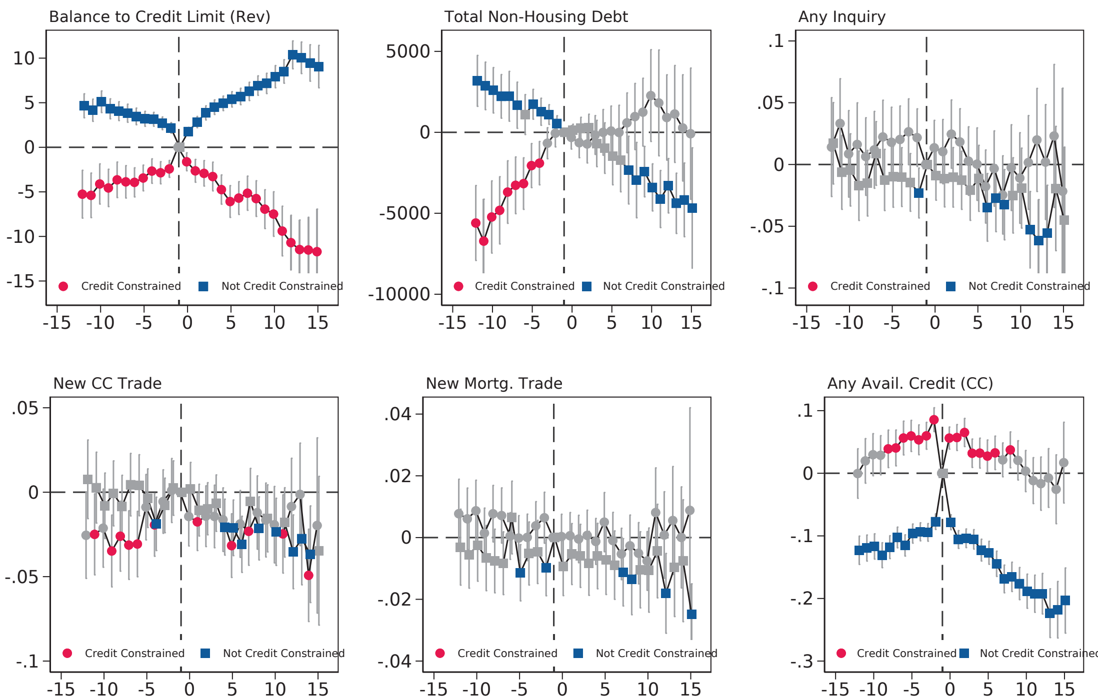

Figure 2.16: Consumption results by credit availability

*Note:*The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

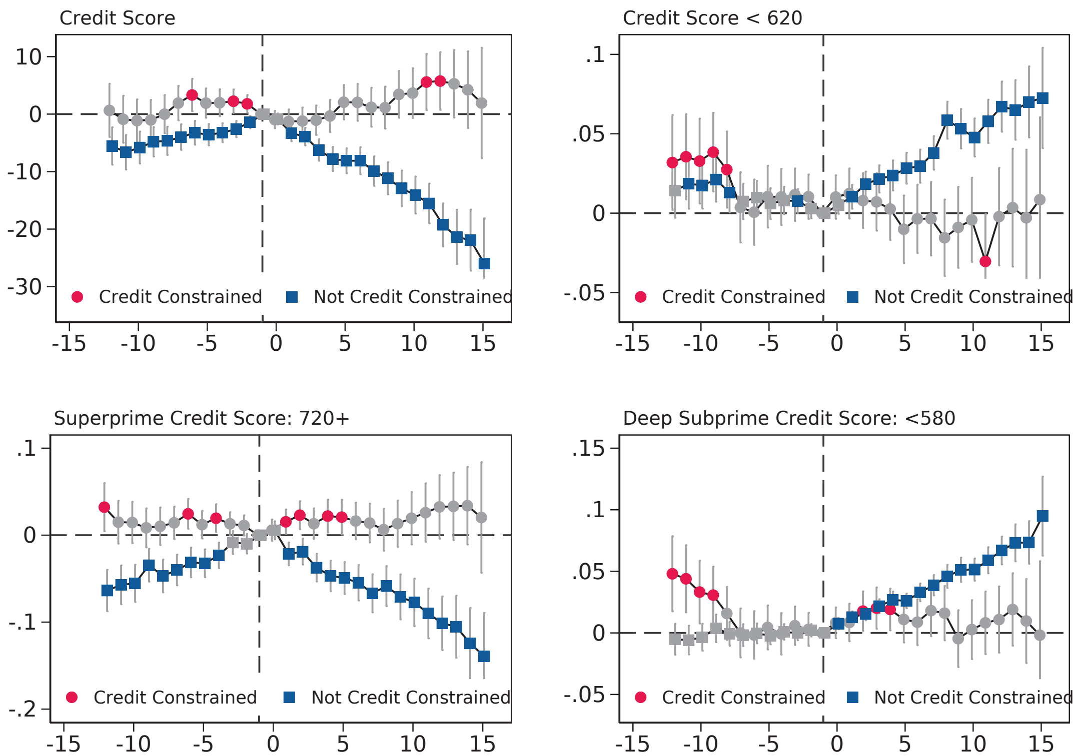

Figure 2.17: Credit performance results by credit availability

*Note:*The results in this figure are congruent to the specification in the main results. An important note is that, while visualized in the same figure, the results displayed are from two separate models and so statistical inference between models should not be made. The figure visualizes the effect of the diabetes diagnosis on the outcome for the plotted group relative to the full set of counterfactuals.

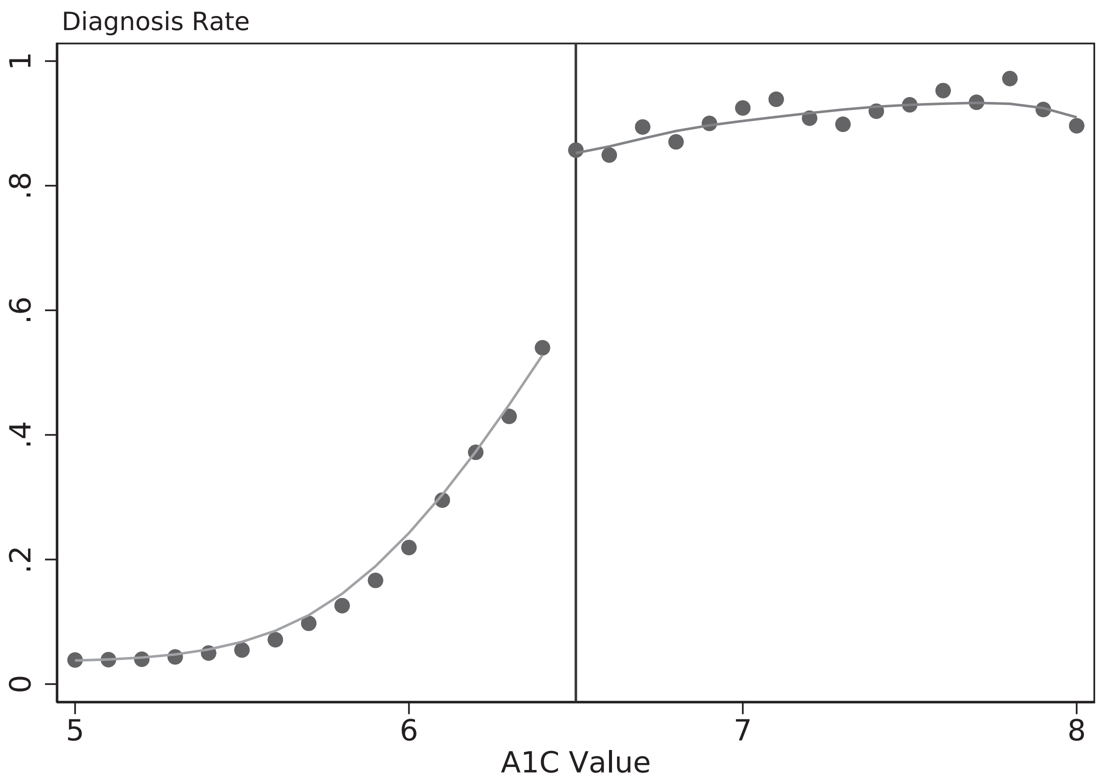

Figure 2.18: Diabetes diagnosis rates around HbA1C threshold
Note: In this figure, diagnosis may occur within four quarters of the quarter of the blood test. It removes individuals who are tested for HbA1C to monitor rather than diagnose. Individuals may appear in single quarter before being diagnosed.

2.5.3 *Fuzzy regression discontinuity design*

The workhorse in RDD estimation in the year of writing is based on the work of Calonico et al. (2014), called "RD Robust".[25] This empirically-driven bandwidth selector, defaults to a linear estimation of the relationship between the running variable and the outcome within an often-narrow bandwidth. Still, the functional form of the relationship is still a meaningful choice both for model fit and for the statistical power that derives from the corresponding bandwidth selection. I choose a cubic polynomial and allow RD Robust to optimally select the bandwidth. Visually, a cubic polynomial appears to fit the data well, especially around a naive bandwidth of $\pm$ 2%, without some of the peculiar quirks that come from the potential over-fit of the quartic model. In Appendix C, I provide more detail about how I selected the functional form selection, and corresponding bandwidths.

The left panel of Figure 2.19 presents the full relationship between HbA1C value and credit score, with horizontal lines at the unconditional means of either side of the threshold. The right panel presents similar information but for employment rate. We can immediately see that in the full set of data, the unconditional means for those on the right half of the threshold, who are substantially more likely to be diagnosed, is lower than for the observations on the left.

[25] The literature on regression discontinuity has shifted from crude or stylistic approaches to empirically-driven estimators of functional form. I fact, though published in 2024, Alalouf et al. (2024) reference an older estimator apparently from Imbens and Kalyanaraman (2012) that appears to have been depreciated, see `https://economics.harvard.edu/faculty/imbens/files/rd_software_09aug4.pdf` though the authors rigorously check alternative bandwidths and specifications.

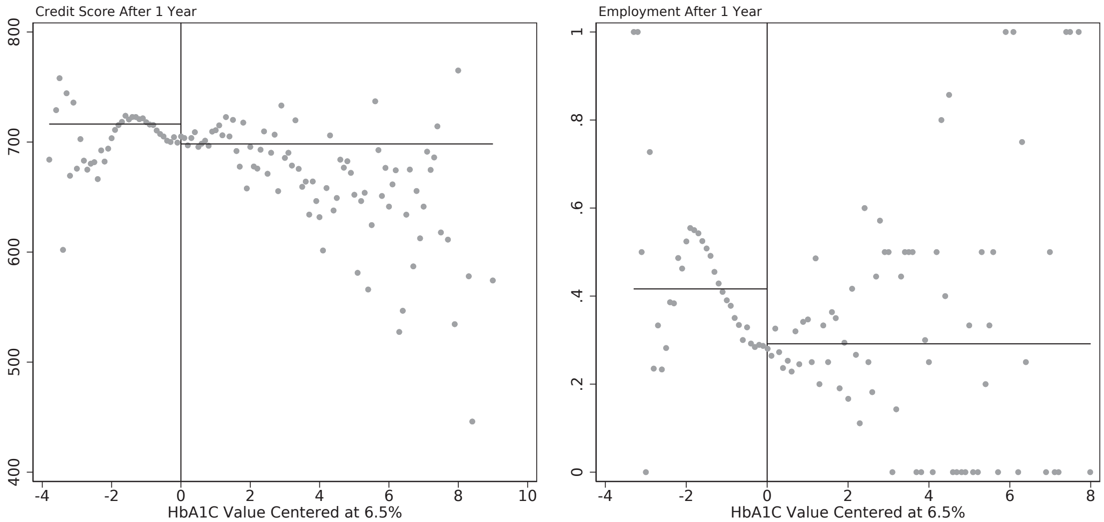

Figure 2.19: Means of credit score and employment rate after one year centered at 6.5% HbA1C

Note: This figure plots the means of the respective outcome within small bins of HbA1C. The figure is centered at 6.5% HbA1C, which is the ADA recommended value for when a patient should be treated for diabetes. The horizontal lines are the unconditional means for the observations to the left and right of the 6.5% threshold.

In Figure 2.20, I zoom in on the relationship to ± 2% of the threshold. In each panel of Figure 2.20, I approximate the relationship with different degrees of polynomials, from linear (p=1) to sextic (p=6).

I present select outcomes here, and the full set of results in the Appendix D for this chapter. In Appendix D, I also try estimates of different functional forms, with different bandwidths for the outcomes presented. The structure of the presentation of the figures is congruent in this section, and is described in the notes of each figure. In short, the focus of the analysis is the colored square in the estimate of the bottom row of the figure.

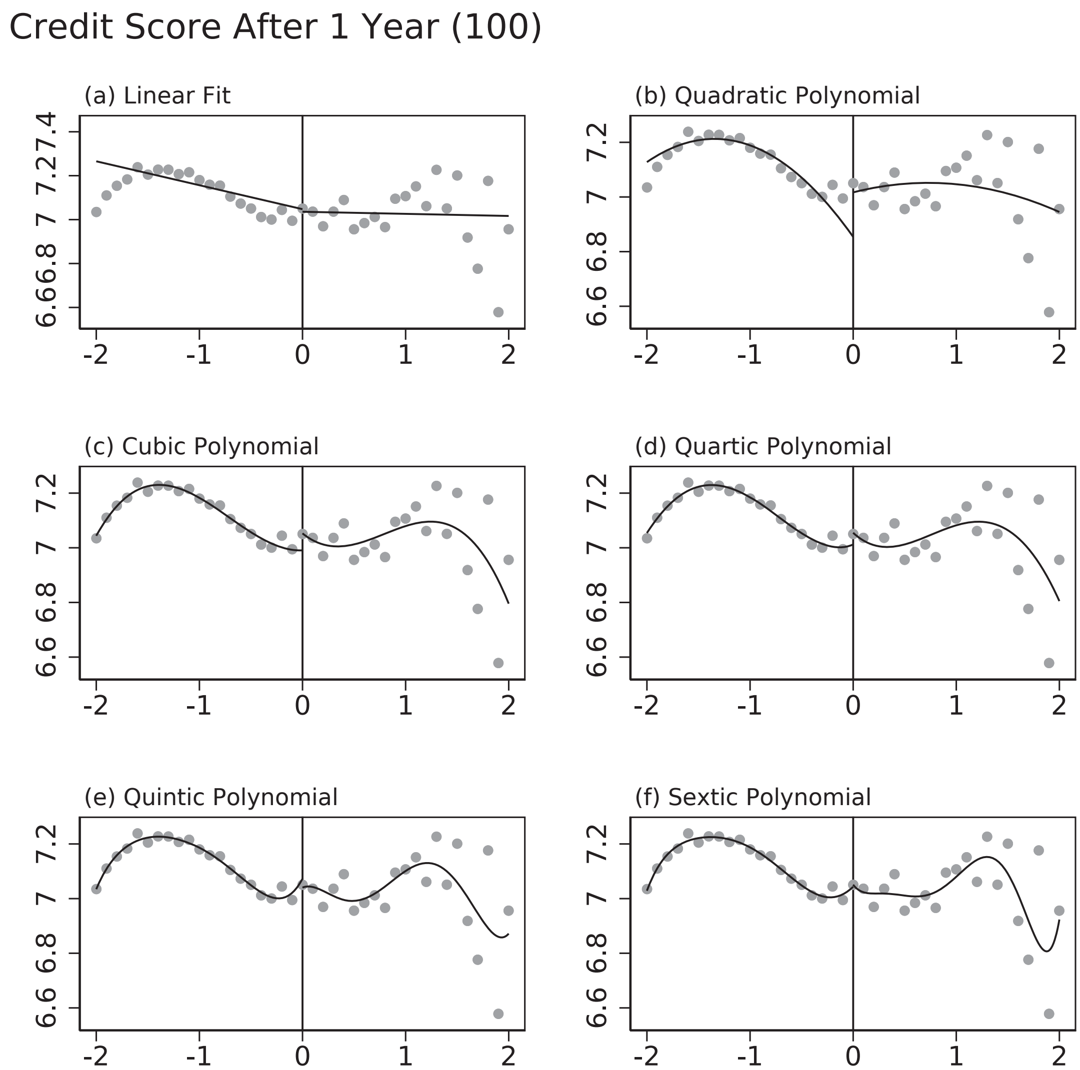

Figure 2.20: Polynomial sensitivity for credit score after one year

Note: This figure zooms in on Figure 2.19 between 6.5% A1C ± 2%. It presents model fits for six different degrees of polynomial functions. Of note is that depending on the polynomial selection, the *a priori* inference may differ. In addition to the polynomial selection, estimates are sensitive to the bandwidth used to construct the polynomial. This characteristic is not displayed in this figure.

Consumption

When analyzing differences in consumption at the margin of diabetes diagnosis, I estimate a statistically-significant decrease in inquiries into new credit lines for those marginally-diagnosed after one year. Figure 2.21 estimates about a 29 percentage point decrease in the likelihood of inquiring for a new line of credit for the marginally-diagnosed, relative to those who were not diagnosed. While inquiries do not necessarily result in new credit lines, I also find that the likelihood of having a new credit trade after one year for the marginally-diagnosed is about 20 percentage points less than those who were not diagnosed, shown in Figure 2.22.

Though not shown, I do not find any statistically significant differences in other consumption outcomes.

Debt Repayment

For those at the margin, there is no difference in having a 60 or more day delinquency in the last year after one to three years of being diagnosed, shown in Figure 2.23. While patients may be spending $1,000 more each year on diabetes-related care (Alalouf et al., 2024), the results of Figure 2.23 suggest that this increased spending is not necessarily resulting in increased delinquency. Indeed, even using different bandwidths and functional forms of estimation (see Appendix D), the inference of this analysis suggest no difference in financial delinquency in the one to three years after diagnosis.

In addition to seeing no difference in the rate of delinquency for patients marginally diagnosed, I also estimate no statistical difference in the amount of medical collections between the two groups in the one to three years after diagnosis. Figure 2.24 provides

Any Inquiry

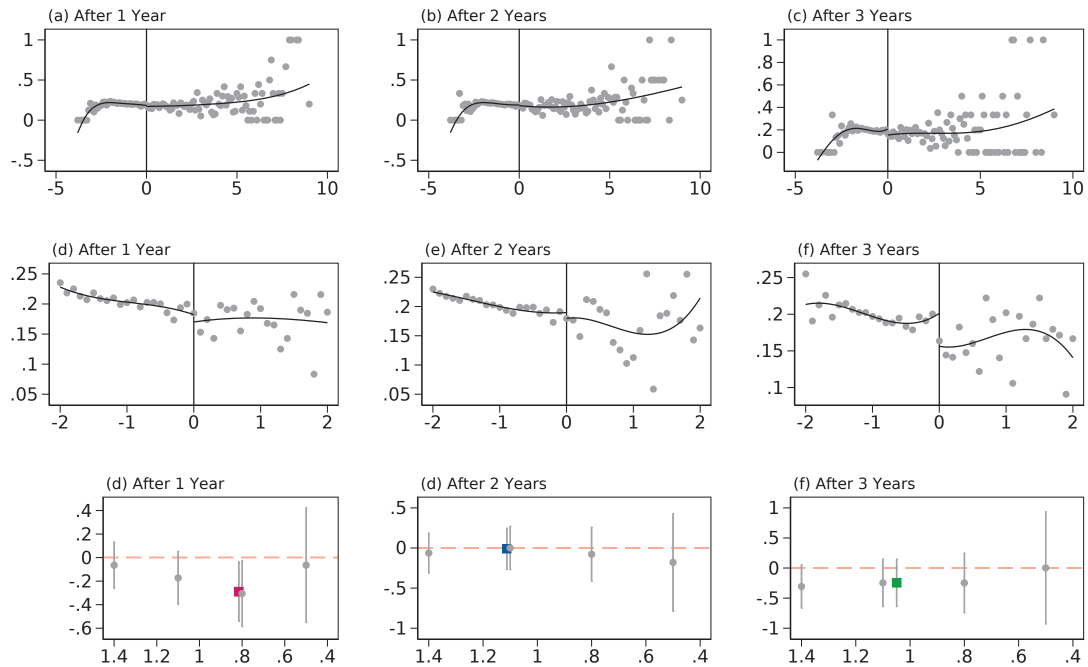

Figure 2.21: Fuzzy RDD results - new credit inquiry

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and two years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

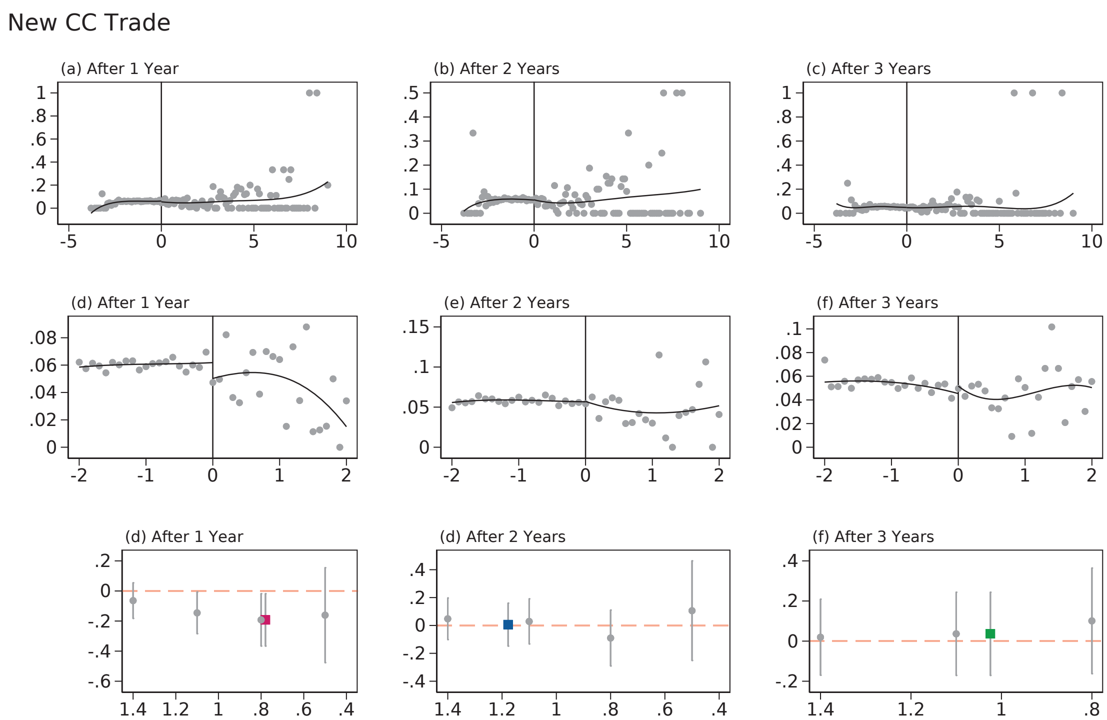

Figure 2.22: Fuzzy RDD results - new CC rrade

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

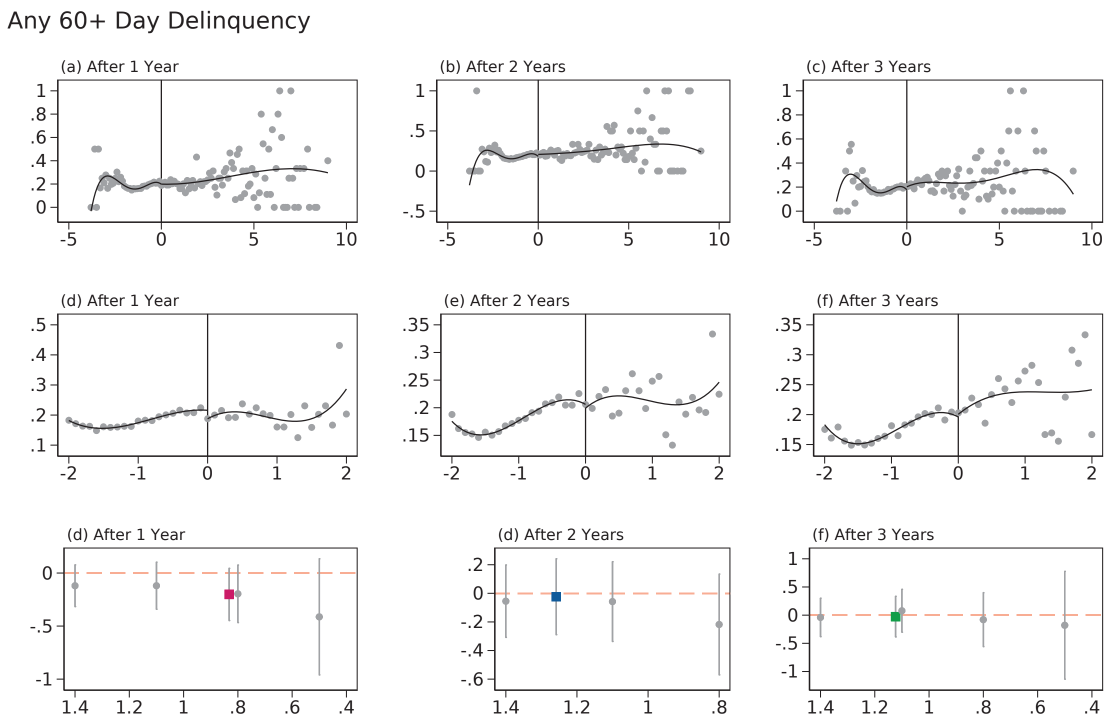

Figure 2.23: Fuzzy RDD results - delinquency

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and two years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

a non-statistically significant increase of a mere $58 in medical collections balance in the first year, and a reduction of $58 in medical collections, relative to those who were not marginally diagnosed. As I discuss below, the framework that I introduced in the 3.2 suggests a trade off between consumption and debt payment, and so the results here fit with this framework when paired with the consumption results above.

I do not show the other outcomes related to debt repayment. I find no statistical difference between the marginally diagnosed and marginally not diagnosed.

Total Medical Collections

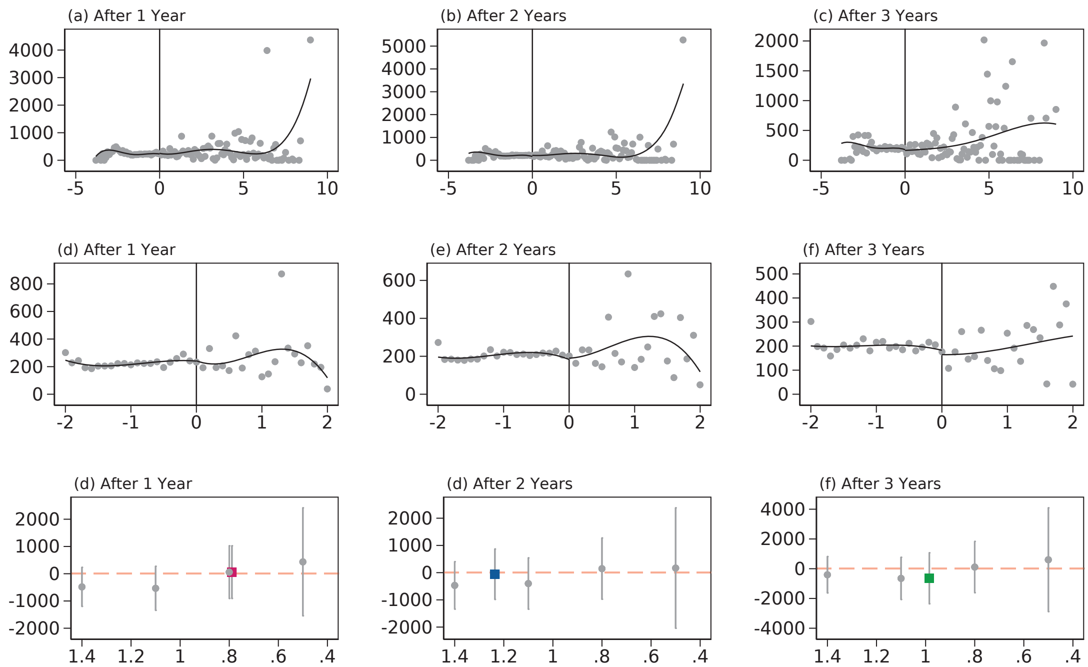

Figure 2.24: Fuzzy RDD results - balance on medical collections

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

Credit performance

Figure 2.19 showed that the unconditional mean credit score for individuals to the right of the threshold was less than the unconditional mean for those to the left. However, the results of the Fuzzy RDD presented in Figure 2.25 suggest that there is no difference in credit score in the one to three years after diagnosis for those diagnosed at the margin. This is not necessarily surprising, since there were no observed differences in debt repayment between the groups. Moreover, in contrast to one of the most prominent results in the difference in differences model, which found a decrease in the likelihood of having a superprime credit score in the quarters of diagnosis, the results of Figure 2.26 find no statistical difference in the likelihood of having a superprime credit score in the one to three years after diagnosis.

Labor supply

Next we turn to labor supply. We saw some visual evidence of potential pre-trends in the event study analysis, and so a regression discontinuity approach here offers an opportunity for further analysis on a group of similar patients. Of note here is that this analysis essentially controls for the difference in physical health between individuals developing diabetes, because it considers individuals at the margin of diagnosis who should otherwise be similar. The only difference between a patient diagnosed with an HbA1C value of 6.6% and one not diagnosed with a value of 6.4% is a small variation in the test value. Making the plausible assumption that the physical ability to supply labor is no different between these two groups, the only difference is the added financial strain of disease management. There are certainly added time costs of managing the disease, and so a decrease in labor supply could reflect this.

Credit Score

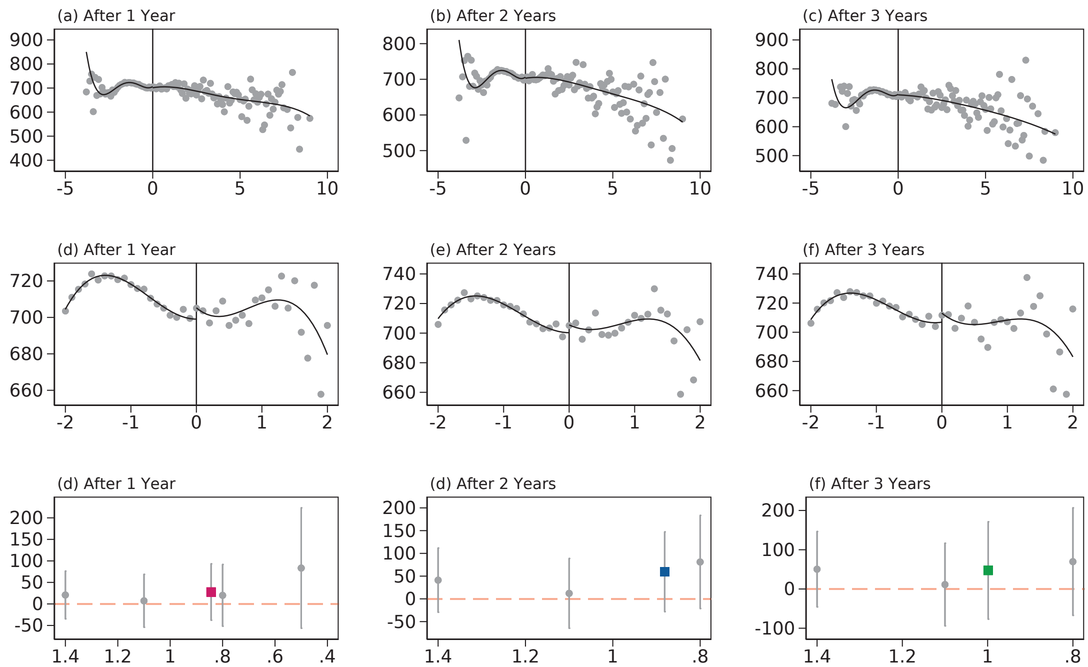

Figure 2.25: Fuzzy RDD results - credit score

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

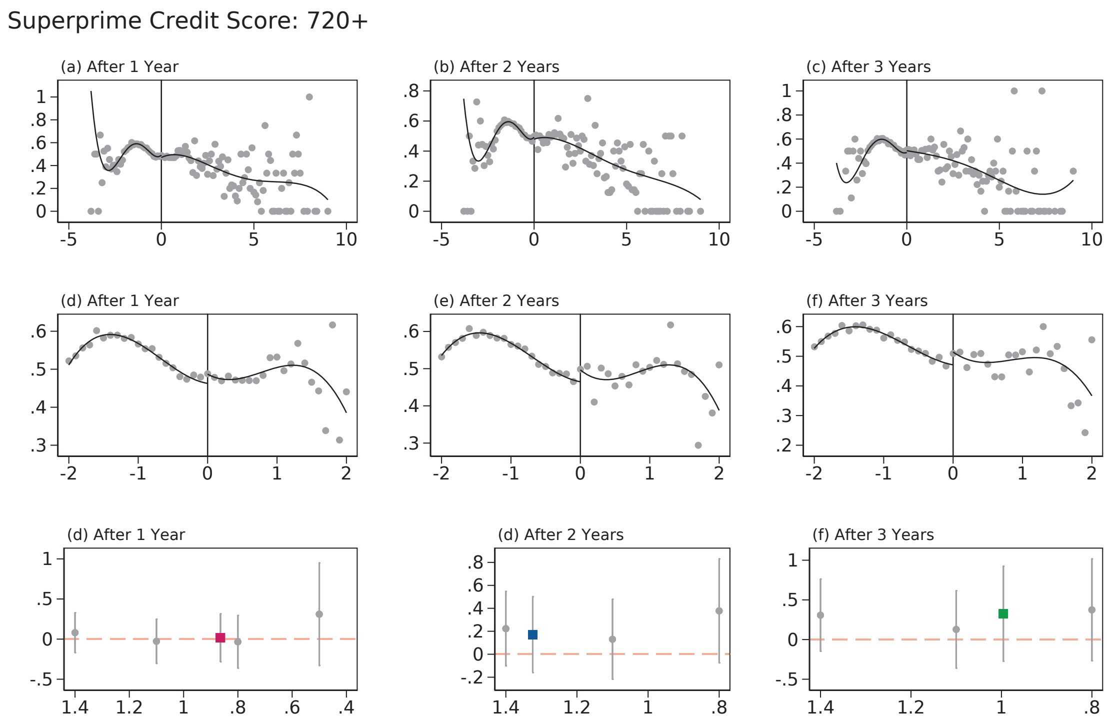

Figure 2.26: Fuzzy RDD results - superprime credit score

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

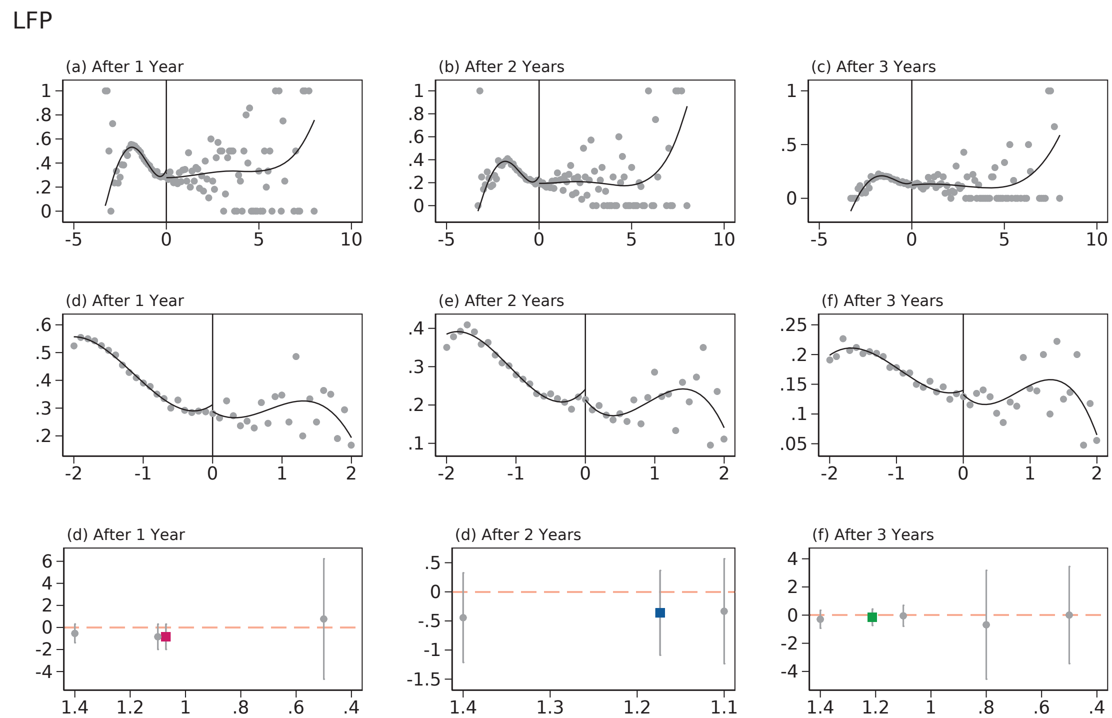

Figure 2.27: Fuzzy RDD results - employment rate

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

Figure 2.27 shows that there is no difference in extensive labor responses for the marginally diagnosed relative to the marginally not diagnosed. Figures 2.28 and 2.29 documents the same for wages and weeks worked, respectively.

Taken together, the results in this section suggest that the consumption response is greatest for those marginally diagnosed. For those on the margin of diagnosis, the added $1,000 per year in medical costs that Alalouf et al. (2024) find appears to result in decreased consumption rather than delinquent debt payments. These two

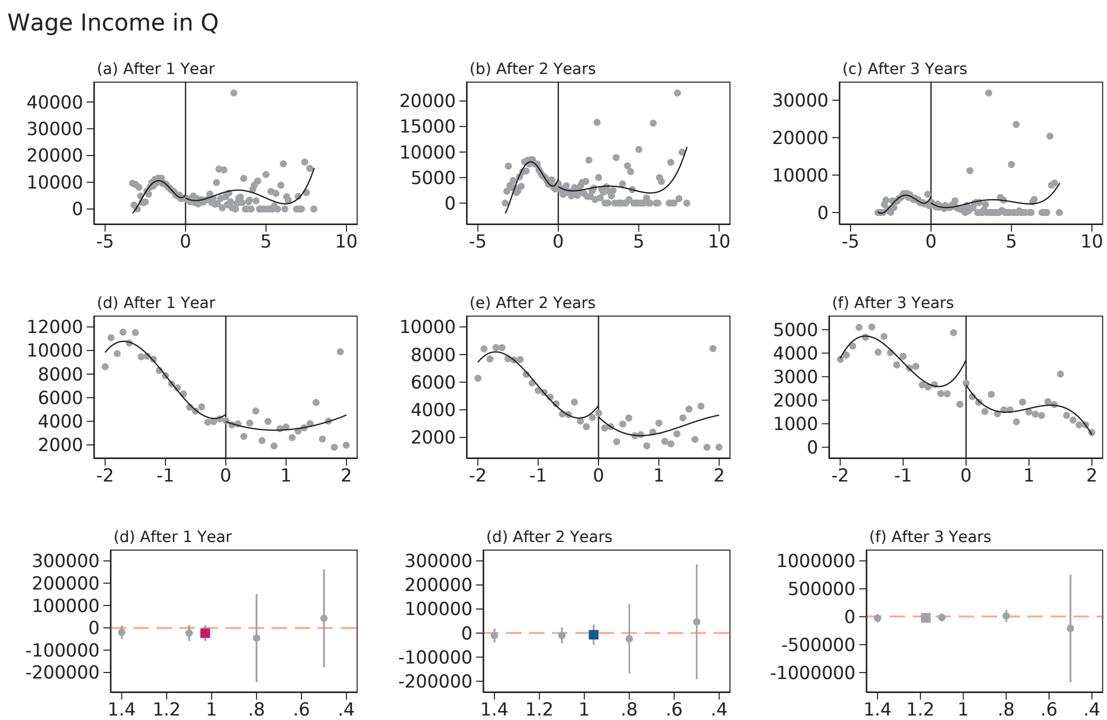

Figure 2.28: Fuzzy RDD results - wages

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

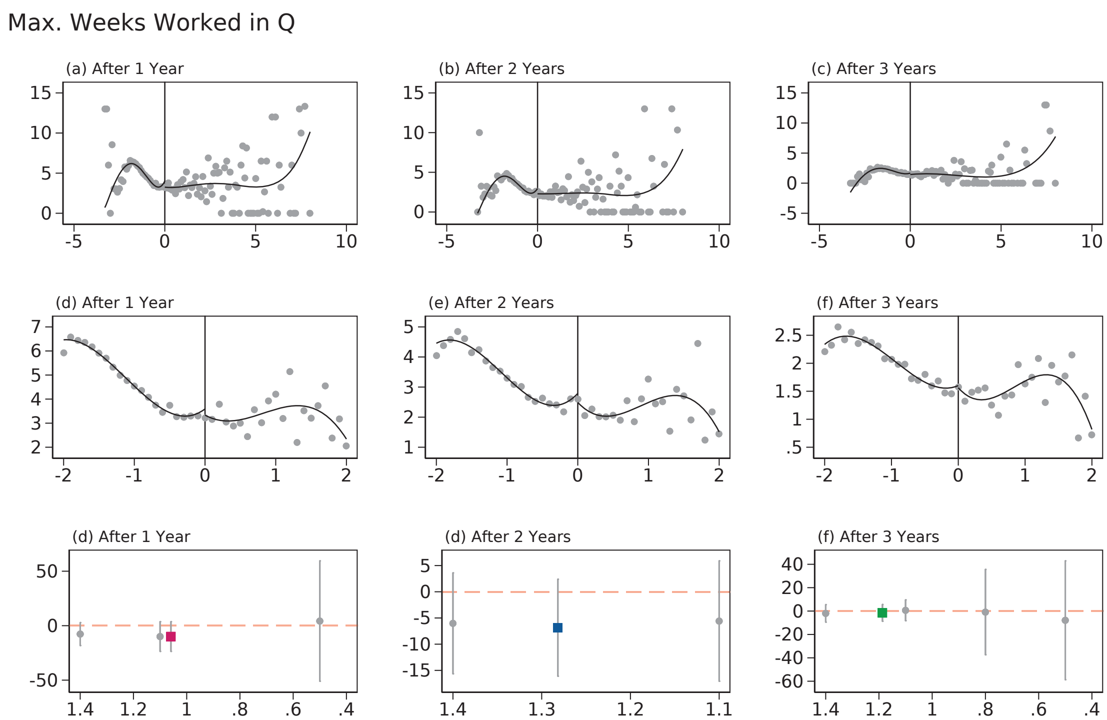

Figure 2.29: Fuzzy RDD results - weeks of labor

Note: The top row presents the full set of data with a quartic line of best fit. Each column presents outcomes as one year, two years, and three years after the HbA1C test. The second row presents the same information as the first row, but zoomed into the bandwidth of ± 2%, and uses the cubic polynomial that I chose to estimate the effect in the bottom row. Row three presents the results of the analysis using a cubic estimator at different bandwidths. The colored square represents the empirically-selected bandwidth, and is the focus of the analysis. The grey circles represent up to five other bandwidths between 0.2% and 1.4% from the 6.5% HbA1C threshold, where estimable.

facts, put together, are strengthened in that I find no statistically significant changes in credit scores. Moreover, the time constraints for the marginally diagnosed are elevated relative to those just below the diagnosis threshold, and so negative point estimates in labor supply outcomes (relative to marginally not diagnosed) make sense. Due to the small sample size at the margin, I do not assess heterogeneity in the RDD results.

2.6 Conclusion

This essay is to my knowledge the first to use causal inference methods to assess the effects of type II diabetes diagnosis on credit outcomes. It is also amongst the first to causally assess labor responses to chronic disease onset, incorporating administrative data on both health and labor supply into econometric models that allow for causal inference. I begin this essay by first conceptualizing how a medical disease might impact the economic behavior of a rational consumer, and then select a handful of attributes from credit and labor data that align with constructs well. In particular, I argue that increased medical costs could manifest in changes to consumption, repayment history, and labor supply. I also make use of credit score as both an aggregated measure for financial health, composed by a number of measures that purport to measure the consumer's ability to repay credit, and an indicator for cost of credit.

I find small impacts to a consumer's debt repayment that result in decreased credit scores from diabetes diagnosis. In particular, I find a precisely-estimated, sustained decrease of only about 2-4 points to a consumer's credit score. Moreover, this decrease in credit score does not appear to be associated with a decrease in credit performance, which can affect the likelihood of obtaining future credit. Consumers are no more likely to have below prime credit scores, but I show evidence that the cost of credit may increase: a decrease of about 6% in the probability of having a superprime credit score suggests that individuals can likely still access credit, though at an increased cost than pre-diagnosis.

At the same time, I find evidence of decreased credit utilization, but little other evidence that consumption changes on aggregate following diagnosis, suggesting that patients on the margin are financing increased cost through inter-temporal substitutions rather than by changes in consumption.

These results appear to be driven by older patients. There is evidence of an increase in the cost of credit for older adults, driven by large increases in delinquency rates and decreases in credit availability. For policymakers, older adults appear to be a group of patients most at risk for financial distress.

Perhaps unique to the medical context, I find pronounced evidence that labor supply decreases as medical costs increase, particularly for single adult households, who presumably have the greatest ability to adjust labor supply. This finding is consistent with the added worker effect (Lundberg, 1985). Whereas individuals may increase their labor supply to offset unexpected costs in typical circumstances, causal techniques in this essay lend support to the literature that has descriptively found negative associations between diabetes and labor supply. In particular, I find a decline of approximately 13% in labor force participation relative to pre-diagnosis. This decline is sustained for at least three years after diagnosis.

One notable exception in the heterogeneity in results by age is that labor responses appear to be driven by the younger group of patients, who perhaps have a greater ability to moderate labor than older adults, who are about 40% less likely to be in the labor force in my data at diagnosis than younger individuals anyway. I explore heterogeneity in the main results, and find that being a member of a household appears to buffer a decline in labor supply, perhaps because of the time costs that are known to be consequences of chronic disease management.

Finally, I also employ a second identification strategy, namely a fuzzy regression discontinuity design, that exploits a HbA1C diagnosis threshold proposed by the American Diabetes Association. This style of modeling offers estimates of a different effect, the local average treatment effect. In contrast to my event study difference in differences results, here I find no statistical difference in the debt repayment at the margin. I find decreases in consumption for the marginally diagnosed, on the other hand. Putting these two findings together suggests that the added costs of type II diabetes diagnosis may be adequately absorbed by changes in consumption rather than by changes in debt payment behavior. Bolstering this finding is that I find no differences in credit performance, which is unsurprising since the components of the credit score were not significantly different between the marginally diagnosed and not diagnosed. On the other hand, when considering the entire diabetes population, it appears that added costs (financial and time) may trigger changes in debt repayment behavior rather than in consumption alone. The fuzzy RDD measures differences only due to changes in disease status (i.e. diagnosed but otherwise the same level of health), whereas the event study in part captures disease progression.

Accordingly, results from this essay point to the notion that financial distress arises from the disease, rather from the diagnosis. While this offers different policy implications (in particular, it provides evidence against simply moving the diagnostic threshold as a way to reduce financial burden), in many ways the implications of these findings are intuitive: reducing the actual prevalence of diabetes, and not just the categorization of diabetes, reduces the economic strain on the American population and allows individuals to remain in the labor force longer. Even then, the results from this essay suggest that the economic strain of diabetes diagnosis is quite minor in aggregate, though certain groups like the elderly and those living alone may experience the strongest adverse effects.

Appendix A: Corresponding tables to diabetes diagnosis event study models, main results, Chapter 2

	Credit Score	Below Prime	Deep Subprime	Subprime	Near Prime	Prime	Superprime
Diag - 16	-3.523	0.023	0.003	0.020	0.00363	-0.04895	-0.00338
	2.567	0.017	0.017	0.016	0.015	0.017674	0.014171
Diag - 15	-1.731	0.026	0.001	0.024	-0.00526	-0.0503	0.010276
	1.844	0.011	0.012	0.011	0.010755	0.012349	0.010036
Diag - 14	-1.990	0.020	0.009	0.011	-0.01605	-0.02552	0.003053
	1.538	0.009	0.010	0.009	0.009115	0.010952	0.008763
Diag - 13	-1.793	0.028	0.017	0.012	-0.02076	-0.02284	0.001406
	1.374	0.009	0.009	0.009	0.00774	0.0096	0.007684
Diag - 12	-0.952	0.019	0.016	0.003	-0.00968	-0.02039	-0.00136
	1.244	0.008	0.008	0.008	0.007522	0.008918	0.007206
Diag - 11	-1.869	0.020	0.014	0.005	-0.00037	-0.02516	-0.0037
	1.126	0.007	0.007	0.007	0.007108	0.008238	0.006628
Diag - 10	-1.679	0.016	0.012	0.004	-0.00249	-0.01844	-0.00443
	1.034	0.007	0.007	0.007	0.006667	0.007992	0.006339
Diag - 9	-1.770	0.024	0.017	0.007	-0.01181	-0.02111	0.001464
	0.966	0.006	0.006	0.006	0.006136	0.007233	0.005767
Diag - 8	-1.080	0.016	0.007	0.009	-0.00787	-0.01142	-0.00448
	0.908	0.006	0.006	0.006	0.006012	0.007034	0.005632
Diag - 7	-0.419	0.006	0.005	0.001	-0.00676	-0.00279	-0.00186
	0.828	0.005	0.005	0.006	0.005812	0.006642	0.005095
Diag - 6	-0.173	0.006	0.008	-0.002	-0.00803	-0.00638	0.00363
	0.772	0.005	0.005	0.005	0.005605	0.00631	0.004808
Diag - 5	-1.400	0.012	0.010	0.002	-0.01112	-0.0008	-0.00346
	0.696	0.005	0.005	0.005	0.005342	0.005987	0.004449
Diag - 4	-1.256	0.014	0.007	0.007	-0.01167	-0.00647	0.001407
	0.623	0.004	0.004	0.005	0.004998	0.005467	0.004235
Diag - 3	-0.802	0.013	0.008	0.005	-0.01268	-0.00497	0.004075
	0.541	0.004	0.004	0.005	0.004627	0.00487	0.003658
Diag - 2	-0.533	0.007	0.007	0.000	-0.01225	0.003865	0.000621
	0.405	0.003	0.003	0.004	0.003919	0.004274	0.003231
Diag + 0	0.035	0.003	0.001	0.003	-0.00454	-0.00245	0.0032
	0.405	0.003	0.003	0.004	0.004006	0.004362	0.003235
Diag + 1	-1.489	0.006	0.002	0.004	0.000111	-0.00471	-0.00603
	0.562	0.004	0.004	0.004	0.004473	0.004935	0.003855
Diag + 2	-1.483	0.005	0.001	0.004	0.00096	-0.01135	-0.00274
	0.627	0.004	0.004	0.005	0.004898	0.005363	0.004318
Diag + 3	-1.857	0.003	0.002	0.001	-0.00246	-0.00089	-0.01142
	0.701	0.005	0.005	0.005	0.00517	0.005895	0.004686
Diag + 4	-2.304	0.004	0.004	0.000	0.002446	-0.0103	-0.01053
	0.772	0.005	0.005	0.005	0.005514	0.006203	0.005045
Diag + 5	-1.855	-0.001	0.000	-0.001	-0.00299	-0.00103	-0.01339
	0.824	0.005	0.005	0.005	0.005715	0.006434	0.005263
Diag + 6	-2.165	0.002	0.002	0.000	-0.0014	-0.00344	-0.01634
	0.878	0.006	0.005	0.006	0.005936	0.006891	0.005664
Diag + 7	-2.124	0.000	0.001	-0.001	-9.3E-05	0.001216	-0.02222
	0.924	0.006	0.006	0.006	0.006201	0.007049	0.005981
Diag + 8	-2.418	-0.001	0.001	-0.002	-0.0009	0.001167	-0.02267
	0.995	0.006	0.006	0.006	0.006336	0.007347	0.006335
Diag + 9	-1.959	-0.001	-0.003	0.001	-0.00366	0.00494	-0.02571
	1.075	0.006	0.006	0.006	0.006515	0.007869	0.006955
Diag + 10	-1.893	0.000	-0.003	0.003	-0.00418	0.003174	-0.02657
	1.122	0.007	0.006	0.007	0.006996	0.00838	0.00742
Diag + 11	-1.491	-0.006	-0.005	-0.001	0.004525	0.00126	-0.02813
	1.204	0.007	0.007	0.007	0.007461	0.008891	0.007893
Diag + 12	-3.122	0.011	0.000	0.011	-0.00273	-0.00826	-0.02874
	1.251	0.007	0.007	0.008	0.007681	0.009388	0.008358
Diag + 13	-3.881	0.008	0.002	0.006	0.003074	-0.01071	-0.03179
	1.364	0.008	0.008	0.008	0.008347	0.010057	0.008772
Diag + 14	-4.302	0.006	-0.002	0.008	0.009965	-0.01684	-0.03522
	1.503	0.009	0.009	0.008	0.009106	0.011127	0.009621
Diag + 15	-4.448	0.009	0.002	0.007	0.010461	-0.01898	-0.0389
	1.709	0.010	0.009	0.009	0.010256	0.01253	0.010836
Diag + 16	-4.274	0.005	-0.003	0.008	0.003445	-0.01224	-0.03958
	1.824	0.011	0.010	0.011	0.011606	0.014115	0.012475
Diag + 17	-6.984	0.004	-0.005	0.010	0.03227	-0.03681	-0.04561
	2.300	0.014	0.014	0.014	0.015594	0.016982	0.014494
Diag + 18	-4.351	-0.022	-0.027	0.005	0.028565	0.003118	-0.07419
	4.292	0.030	0.028	0.026	0.031558	0.029504	0.021654
N	1761937.000	1792997.000	1792997.000	1792997.000	1792997	1792997	1792997
DV Mean	704.456	0.262	0.194	0.068	0.075114	0.133598	0.512144
Baseline DV Mean	705.264	0.260	0.193	0.067	0.07423	0.132852	0.51614

Table A.1: Tabular results of credit performance models

	Individual Debt	CC Bal to Credit	Rev. Bal to Credit	New CC Trade	New Auto Trade	New Personal Trade	New Mort Trade	New Stu Trade	Any CC Credit Avail	Any Rev Credit Avail	Housing Debt	Non-Housing Deb	New Inquiry
Diag - 16	4404.522	-0.545	0.851	-0.026	0.014	0.007	0.002	0.015	-0.012	0.057	3367.601	1028.237	0.035
	4074.148	1.673	1.610	0.012	0.012	0.006	0.007	0.008	0.018	0.029	3730.018	1226.444	0.023
Diag - 15	3653.390	1.506	1.502	-0.020	0.014	-0.002	-0.001	-0.003	-0.028	0.018	3737.585	4.118	0.009
	2538.807	1.157	1.063	0.009	0.008	0.002	0.004	0.003	0.014	0.022	2333.105	861.599	0.016
Diag - 14	4588.828	1.839	0.998	-0.006	-0.002	-0.002	0.003	0.001	-0.027	0.018	4852.274	-273.537	-0.002
	2021.800	0.991	0.903	0.008	0.006	0.002	0.004	0.003	0.011	0.018	1863.938	728.758	0.014
Diag - 13	4550.213	1.249	1.043	-0.010	-0.006	-0.001	0.001	0.000	-0.018	0.037	5219.172	-600.654	0.014
	1791.121	0.892	0.779	0.007	0.005	0.002	0.003	0.003	0.010	0.017	1640.963	700.803	0.012
Diag - 12	4467.576	0.631	0.266	-0.003	-0.002	0.003	0.004	-0.001	-0.021	0.029	4721.657	-204.394	0.018
	1629.599	0.815	0.731	0.007	0.005	0.003	0.003	0.002	0.009	0.016	1522.182	588.611	0.011
Diag - 11	3686.473	0.266	0.091	-0.006	-0.001	0.001	0.002	-0.001	-0.021	0.019	4450.474	-702.831	0.019
	1583.094	0.743	0.674	0.006	0.005	0.002	0.003	0.002	0.009	0.015	1482.667	587.609	0.010
Diag - 10	4283.163	1.373	0.793	-0.010	-0.003	0.002	0.004	-0.001	-0.010	0.027	4810.657	-498.821	0.009
	1348.815	0.689	0.627	0.006	0.004	0.002	0.003	0.002	0.008	0.014	1242.428	523.604	0.010
Diag - 9	2225.209	0.933	0.278	-0.010	-0.004	0.000	0.000	0.000	-0.017	-0.010	2764.141	-539.057	-0.001
	1276.539	0.638	0.583	0.006	0.004	0.002	0.003	0.002	0.008	0.014	1168.820	489.373	0.009
Diag - 8	2781.270	0.789	0.589	-0.011	-0.005	0.000	0.001	0.000	-0.012	0.004	2961.289	-184.892	-0.004
	1193.935	0.589	0.554	0.005	0.004	0.002	0.003	0.002	0.007	0.013	1068.252	489.288	0.009
Diag - 7	1517.231	0.219	0.362	-0.007	-0.001	-0.001	0.001	-0.002	-0.009	0.014	1999.767	-421.746	0.010
	1124.312	0.568	0.516	0.005	0.004	0.002	0.003	0.002	0.007	0.012	1016.190	450.691	0.009
Diag - 6	928.855	0.362	0.204	-0.008	-0.001	0.001	0.004	-0.002	-0.011	-0.005	1738.163	-788.122	0.010
	1006.527	0.535	0.476	0.005	0.004	0.002	0.003	0.002	0.006	0.012	905.865	404.636	0.008
Diag - 5	520.741	0.714	0.593	-0.002	0.007	0.001	-0.003	-0.002	-0.005	0.014	712.224	-193.422	0.011
	884.484	0.485	0.433	0.005	0.004	0.002	0.002	0.002	0.006	0.011	808.097	337.185	0.008
Diag - 4	-68.845	0.212	0.637	-0.013	-0.002	0.001	-0.001	-0.002	-0.012	-0.005	250.084	-327.842	0.004
	809.048	0.443	0.402	0.005	0.004	0.002	0.002	0.002	0.006	0.010	744.190	318.283	0.008
Diag - 3	721.497	0.072	0.396	-0.003	-0.002	0.000	0.000	-0.001	-0.012	0.007	652.406	54.917	0.006
	687.857	0.403	0.367	0.005	0.004	0.002	0.003	0.002	0.005	0.009	640.618	244.785	0.008
Diag - 2	762.402	0.622	0.465	0.002	-0.004	0.000	-0.002	-0.001	-0.003	0.019	636.577	101.294	0.001
	523.817	0.339	0.301	0.005	0.004	0.002	0.002	0.002	0.005	0.009	481.482	201.015	0.007
Diag + 0	-652.226	-0.130	-0.029	0.003	-0.002	0.000	-0.003	0.000	-0.008	-0.012	-594.549	-59.145	0.007
	479.023	0.329	0.291	0.005	0.004	0.002	0.002	0.002	0.004	0.008	442.020	175.944	0.006
Diag + 1	-1460.567	-0.048	0.323	0.001	-0.001	0.001	-0.002	0.000	-0.009	-0.015	-1417.264	-54.683	0.004
	686.383	0.401	0.364	0.004	0.004	0.002	0.002	0.001	0.005	0.010	634.800	250.266	0.007
Diag + 2	-2223.058	0.431	0.415	-0.001	-0.003	0.004	-0.003	0.000	-0.003	-0.016	-2184.279	-56.352	0.005
	825.179	0.446	0.407	0.005	0.004	0.002	0.002	0.002	0.006	0.010	760.051	300.885	0.008
Diag + 3	-3760.384	0.148	0.201	0.002	0.003	0.001	-0.005	0.000	-0.010	-0.037	-3461.235	-300.614	0.006
	1018.696	0.475	0.449	0.005	0.004	0.002	0.002	0.002	0.006	0.010	895.273	382.652	0.007
Diag + 4	-3736.185	0.185	-0.106	0.001	-0.004	0.003	-0.002	0.000	-0.007	-0.055	-3306.159	-420.605	0.002
	1192.566	0.509	0.462	0.005	0.004	0.002	0.002	0.002	0.006	0.012	1058.208	427.696	0.008
Diag + 5	-3807.717	-0.054	-0.276	-0.005	-0.002	0.000	-0.003	0.000	-0.014	-0.064	-3527.778	-279.368	0.007
	1236.067	0.546	0.496	0.005	0.004	0.002	0.003	0.002	0.007	0.012	1098.708	456.092	0.008
Diag + 6	-4583.537	0.766	0.392	-0.003	-0.003	0.001	-0.005	-0.002	-0.004	-0.064	-4154.393	-438.085	-0.007
	1265.900	0.574	0.510	0.005	0.004	0.002	0.003	0.002	0.007	0.013	1098.273	496.342	0.008
Diag + 7	-5350.613	0.657	0.440	-0.002	-0.002	0.001	-0.008	-0.001	-0.012	-0.089	-4847.322	-501.055	-0.003
	1332.382	0.592	0.532	0.005	0.004	0.002	0.002	0.002	0.007	0.014	1150.757	528.206	0.008
Diag + 8	-5744.211	0.679	0.352	-0.001	-0.001	0.003	-0.007	-0.003	-0.014	-0.094	-5415.858	-353.420	-0.003
	1415.443	0.617	0.557	0.005	0.004	0.002	0.003	0.002	0.008	0.014	1203.196	578.037	0.009
Diag + 9	-6837.229	0.447	0.104	0.002	0.005	0.003	-0.009	-0.001	-0.016	-0.100	-6535.659	-361.363	0.004
	1329.221	0.653	0.596	0.006	0.004	0.003	0.003	0.002	0.008	0.015	1182.191	506.142	0.009
Diag + 10	-7912.331	0.559	0.242	-0.003	-0.003	0.002	-0.008	-0.001	-0.016	-0.113	-7918.472	-65.353	-0.005
	1356.607	0.723	0.649	0.006	0.004	0.003	0.003	0.002	0.009	0.015	1231.125	584.579	0.010
Diag + 11	-9467.179	0.039	-0.349	0.003	-0.003	0.000	-0.002	-0.001	-0.023	-0.130	-8798.073	-735.124	-0.012
	1513.119	0.769	0.698	0.006	0.004	0.003	0.003	0.002	0.009	0.015	1379.957	640.025	0.010
Diag + 12	-9957.768	0.000	-0.195	0.001	-0.003	0.006	-0.011	-0.003	-0.024	-0.135	-9239.388	-798.801	0.000
	1500.560	0.832	0.759	0.006	0.005	0.003	0.003	0.002	0.010	0.017	1335.957	589.760	0.010
Diag + 13	-10766.261	-0.396	0.206	0.004	-0.001	-0.002	-0.005	-0.002	-0.027	-0.140	-9755.939	-1090.988	-0.018
	1664.554	0.863	0.774	0.007	0.005	0.003	0.004	0.002	0.011	0.018	1495.537	643.812	0.011
Diag + 14	-12974.208	-0.940	-0.112	-0.010	-0.002	-0.001	-0.008	0.002	-0.025	-0.152	-11857.605	-1237.996	0.002
	1787.812	0.937	0.843	0.007	0.005	0.003	0.004	0.003	0.011	0.019	1591.230	689.995	0.012
Diag + 15	-13459.483	-1.051	-1.221	0.010	0.007	-0.003	-0.005	0.002	-0.036	-0.171	-11814.810	-1782.759	-0.018
	1921.033	1.018	0.922	0.009	0.007	0.003	0.005	0.003	0.012	0.020	1722.601	760.057	0.013
Diag + 16	-13967.836	-1.816	-2.406	-0.007	-0.004	0.005	-0.009	0.000	-0.018	-0.166	-11898.920	-2252.930	-0.008
	2020.542	1.124	1.011	0.009	0.007	0.005	0.005	0.003	0.014	0.021	1807.225	822.700	0.015
Diag + 17	-16199.671	-0.674	-0.669	-0.012	-0.012	-0.001	0.002	0.001	-0.030	-0.185	-13685.567	-2590.615	0.002
	2185.051	1.605	1.423	0.012	0.007	0.005	0.008	0.003	0.018	0.028	1887.388	917.014	0.019
Diag + 18	-18328.598	-4.185	-2.707	-0.015	-0.003	-0.005	-0.007	-0.002	-0.053	-0.189	-15700.888	-2557.915	0.040
	5727.378	3.012	2.873	0.019	0.014	0.011	0.010	0.002	0.028	0.045	5516.655	1284.039	0.034
N	1792997	1244704	1401082	1792997	1792997	1792997	1792997	1792997	1792997	1726023	1792997	1792997	1792997
DV Mean	86748.249	29.349	26.858	0.060	0.032	0.009	0.018	0.009	0.566	0.678	66006.683	21690.711	0.211
Baseline DV Mean	87968.057	29.098	26.647	0.060	0.032	0.008	0.018	0.010	0.569	0.678	67115.254	21812.427	0.211

Table A.2: Tabular results of consumption models

	Total Collections	Delinquency	Chargeoffs	Dismissed BK	Discharged BK	Foreclosure	Medical Collections
Diag - 16	-121.164	0.028	-0.006	-0.001	0.001	0.002	-52.864
	87.270	0.019	0.013	0.003	0.003	0.003	44.071
Diag - 15	-99.684	0.024	0.000	0.000	-0.003	0.000	-49.153
	70.379	0.013	0.009	0.002	0.001	0.001	47.708
Diag - 14	-48.395	0.023	0.005	-0.003	0.000	0.000	-51.443
	69.506	0.011	0.007	0.001	0.002	0.001	37.956
Diag - 13	-55.462	0.030	0.000	0.000	0.001	0.001	-67.360
	58.802	0.010	0.006	0.002	0.002	0.001	32.964
Diag - 12	14.635	0.015	-0.001	0.000	0.000	-0.001	-27.749
	54.888	0.009	0.006	0.001	0.001	0.000	35.490
Diag - 11	67.252	0.023	0.001	-0.001	0.000	0.000	17.176
	54.535	0.008	0.005	0.001	0.001	0.001	40.410
Diag - 10	63.249	0.023	0.003	-0.001	0.000	0.000	-6.425
	43.861	0.008	0.005	0.001	0.001	0.001	25.769
Diag - 9	20.261	0.020	0.004	0.000	0.001	0.000	-18.882
	39.696	0.007	0.004	0.001	0.001	0.000	23.396
Diag - 8	38.220	0.020	0.001	-0.001	0.000	0.001	-5.496
	37.021	0.007	0.004	0.001	0.001	0.001	22.272
Diag - 7	9.289	0.011	0.003	-0.001	0.001	0.000	-13.025
	34.891	0.007	0.004	0.001	0.001	0.001	21.391
Diag - 6	2.877	0.012	0.001	0.000	0.000	0.000	-16.188
	32.093	0.006	0.003	0.001	0.001	0.001	18.875
Diag - 5	8.032	0.009	0.000	-0.001	0.000	0.000	-7.579
	28.630	0.006	0.003	0.001	0.001	0.000	17.851
Diag - 4	25.191	0.005	0.001	-0.001	0.000	0.000	-2.718
	25.279	0.005	0.003	0.001	0.001	0.000	15.533
Diag - 3	18.814	0.005	0.003	-0.001	0.001	0.000	-11.492
	21.194	0.004	0.002	0.001	0.001	0.000	14.265
Diag - 2	2.264	0.002	0.001	-0.002	0.002	0.000	-3.736
	15.240	0.003	0.001	0.001	0.001	0.000	10.692
Diag + 0	-11.378	0.004	0.002	-0.002	0.002	0.000	-11.607
	16.252	0.003	0.001	0.001	0.001	0.000	8.759
Diag + 1	-11.294	0.012	0.004	-0.002	-0.001	0.000	-0.966
	24.691	0.004	0.002	0.001	0.001	0.000	18.341
Diag + 2	-10.581	0.011	0.003	-0.001	0.000	0.000	-8.651
	27.399	0.005	0.002	0.001	0.001	0.000	20.121
Diag + 3	-20.256	0.012	0.004	0.000	0.000	0.000	-21.642
	25.354	0.005	0.003	0.001	0.001	0.001	14.701
Diag + 4	-30.853	0.013	0.004	-0.001	0.001	0.000	-12.069
	26.449	0.006	0.003	0.001	0.001	0.001	16.312
Diag + 5	-39.229	0.012	0.006	-0.001	0.001	0.000	-18.235
	29.496	0.006	0.003	0.001	0.001	0.001	18.378
Diag + 6	-77.876	0.012	0.007	-0.001	0.000	0.000	-50.520
	30.471	0.007	0.004	0.001	0.001	0.001	17.387
Diag + 7	-60.844	0.013	0.006	0.000	0.000	0.000	-41.865
	33.726	0.007	0.004	0.001	0.001	0.001	18.423
Diag + 8	-104.601	0.014	0.005	0.000	0.002	0.001	-37.871
	39.553	0.007	0.004	0.001	0.001	0.001	21.641
Diag + 9	-137.829	0.016	0.004	0.001	0.001	0.000	-53.941
	47.844	0.007	0.005	0.001	0.001	0.001	31.968
Diag + 10	-122.041	0.007	0.001	0.000	0.002	-0.001	-40.400
	51.462	0.008	0.005	0.001	0.001	0.001	35.121
Diag + 11	-164.692	0.001	0.002	0.000	0.002	0.000	-51.487
	55.640	0.008	0.005	0.001	0.001	0.001	37.534
Diag + 12	-158.314	0.005	0.003	-0.001	0.001	0.000	-43.733
	60.826	0.008	0.005	0.001	0.001	0.001	41.845
Diag + 13	-154.745	0.005	0.001	-0.001	-0.001	0.000	-52.539
	67.993	0.009	0.006	0.001	0.001	0.001	45.541
Diag + 14	-147.892	0.003	0.003	-0.001	0.000	0.000	-62.444
	68.455	0.010	0.006	0.001	0.001	0.001	36.001
Diag + 15	-111.943	0.000	0.003	0.000	0.000	0.000	-22.368
	76.784	0.011	0.006	0.002	0.001	0.001	29.821
Diag + 16	-122.716	-0.001	-0.004	-0.002	-0.001	0.000	-35.279
	92.779	0.012	0.008	0.001	0.001	0.001	29.807
Diag + 17	-42.982	-0.003	-0.011	-0.002	0.002	0.003	-63.219
	77.644	0.015	0.010	0.001	0.003	0.003	37.665
Diag + 18	-85.976	-0.025	-0.010	-0.002	-0.001	0.000	-52.207
	137.159	0.025	0.021	0.001	0.001	0.001	41.356
N	1792997.000	1792997.000	1792997.000	1792997.000	1792997.000	1792997.000	1792997.000
DV Mean	651.795	0.109	0.059	0.001	0.001	0.000	254.999
Baseline DV Mean	646.227	0.108	0.058	0.001	0.001	0.000	253.773

Table A.3: Tabular results of debt repayment models

	Number of Employers	Weeks Worked	Wages	LFP
Diag - 19	-0.052	0.136	690.913	-0.026
	-0.044	-0.389	-510.889	-0.033
Diag - 18	0.010	0.324	1101.158	0.003
	-0.043	-0.329	-429.967	-0.028
Diag - 17	0.014	0.529	1439.684	0.015
	-0.033	-0.290	-573.924	-0.023
Diag - 16	0.042	0.605	882.333	0.015
	0.032	0.274	533.465	0.022
Diag - 15	0.030	0.450	866.581	0.016
	0.026	0.233	482.800	0.019
Diag - 14	0.030	0.479	942.236	0.017
	0.026	0.230	409.702	0.019
Diag - 13	0.024	0.471	1482.570	0.029
	0.023	0.213	450.847	0.018
Diag - 12	-0.003	0.446	1178.038	0.022
	0.020	0.186	431.323	0.015
Diag - 11	0.029	0.512	1073.331	0.026
	0.019	0.174	446.582	0.014
Diag - 10	0.032	0.455	793.259	0.030
	0.020	0.172	386.127	0.013
Diag - 9	0.022	0.342	912.703	0.015
	0.019	0.158	304.128	0.013
Diag - 8	0.023	0.144	312.014	0.011
	0.018	0.150	344.605	0.012
Diag - 7	0.024	0.231	235.499	0.006
	0.017	0.137	368.983	0.011
Diag - 6	0.031	0.276	378.179	0.016
	0.016	0.127	280.961	0.010
Diag - 5	0.021	0.187	381.476	0.012
	0.014	0.117	208.888	0.010
Diag - 4	0.014	0.039	28.657	0.004
	0.014	0.111	290.596	0.009
Diag - 3	0.017	0.124	-4.431	0.005
	0.013	0.094	323.764	0.008
Diag - 2	0.008	-0.010	-120.020	0.004
	0.011	0.081	250.343	0.007
Diag + 0	-0.016	-0.184	-519.856	-0.012
	0.010	0.072	261.472	0.006
Diag + 1	-0.027	-0.332	-751.116	-0.029
	0.012	0.094	326.488	0.008
Diag + 2	-0.018	-0.307	-839.572	-0.027
	0.013	0.107	261.729	0.009
Diag + 3	-0.030	-0.406	-866.678	-0.033
	0.015	0.114	246.663	0.010
Diag + 4	-0.024	-0.329	-570.551	-0.030
	0.016	0.125	465.701	0.011
Diag + 5	-0.051	-0.493	-1099.858	-0.042
	0.016	0.135	305.272	0.011
Diag + 6	-0.049	-0.509	-1209.398	-0.042
	0.017	0.142	322.329	0.012
Diag + 7	-0.027	-0.550	-1195.132	-0.041
	0.021	0.150	312.153	0.013
Diag + 8	-0.066	-0.560	-1405.150	-0.056
	0.020	0.169	425.085	0.014
Diag + 9	-0.054	-0.579	-1497.662	-0.049
	0.021	0.168	395.800	0.014
Diag + 10	-0.063	-0.718	-1646.481	-0.056
	0.021	0.180	369.647	0.015
Diag + 11	-0.073	-0.829	-1983.569	-0.066
	0.023	0.200	407.870	0.016
Diag + 12	-0.062	-0.742	-1614.798	-0.070
	0.024	0.216	531.314	0.018
Diag + 13	-0.047	-0.683	-2060.659	-0.059
	0.027	0.224	500.198	0.019
Diag + 14	-0.040	-0.626	-1890.659	-0.049
	0.027	0.245	568.195	0.021
Diag + 15	0.002	-0.322	-1370.464	-0.016
	0.034	0.278	509.439	0.023
Diag + 16	-0.025	-0.614	-941.247	-0.037
	0.033	0.326	698.620	0.025
Diag + 17	0.041	-0.097	-1001.438	0.019
	0.036	0.350	618.176	0.025
Diag + 18	-0.056	0.090	-1696.012	0.001
	0.056	0.654	775.658	0.046
N	439297.000	439297.000	439297.000	439297.000
DV Mean	0.641	6.402	9807.249	0.552
Baseline DV Mean	0.648	6.473	10002.312	0.558

Table A.4: Tabular results of labor supply models

Appendix B: Additional diabetes event study outcomes, Chapter 2

Figure B.1: Prime credit score (660-719)

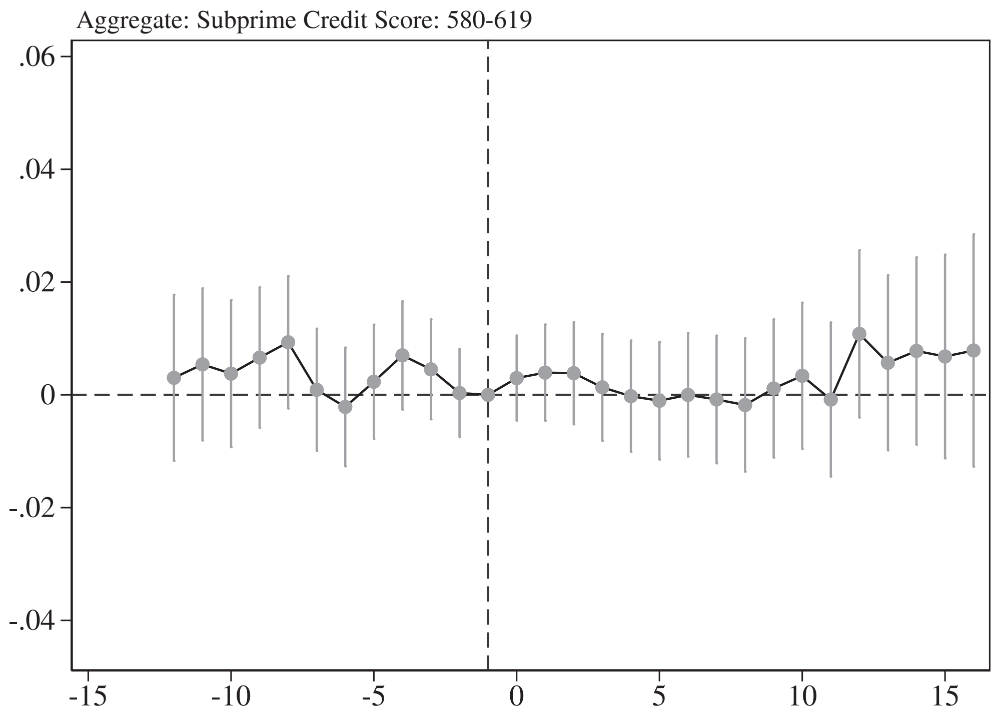

Figure B.2: Subprime credit score (580-619)

Appendix C: Details and analysis on RDD functional form and bandwidth selection, Chapter 2

If we consider the relationship between credit score after 1 year of a diabetes test centered on distance away from the HbA1C threshold, the choice of functional form and of bandwidth becomes apparent. I choose this example because credit score can be viewed as a summary measure of financial health, and because I found earlier that diabetes diagnosis reduces credit score by around 2 points in the 2 years after diagnosis.

Complicating the RDD model selection further is the interaction of bandwidth selection. A very narrow bandwidth will be required to approximate any sort of linear relationship between these variables. As a result, few observations may actually fall in that bandwidth, and the statistical power may be too low to observe a real effect. RD Robust selects an optimal bandwidth for a given polynomial. Figure C.1 uses a quartic (p=4) polynomial for data of different bandwidths. In panel (a), the full domain of data is used, and the domain shrinks until panel (f) which uses the domain of $[-0.5, 0.5]$. As the bandwidth widens, it is apparent that a quartic polynomial might be overfitting the data. There is certainly a rational argument that could be made for a narrow bandwidth of data used for a first-degree polynomial, the default in RD Robust.

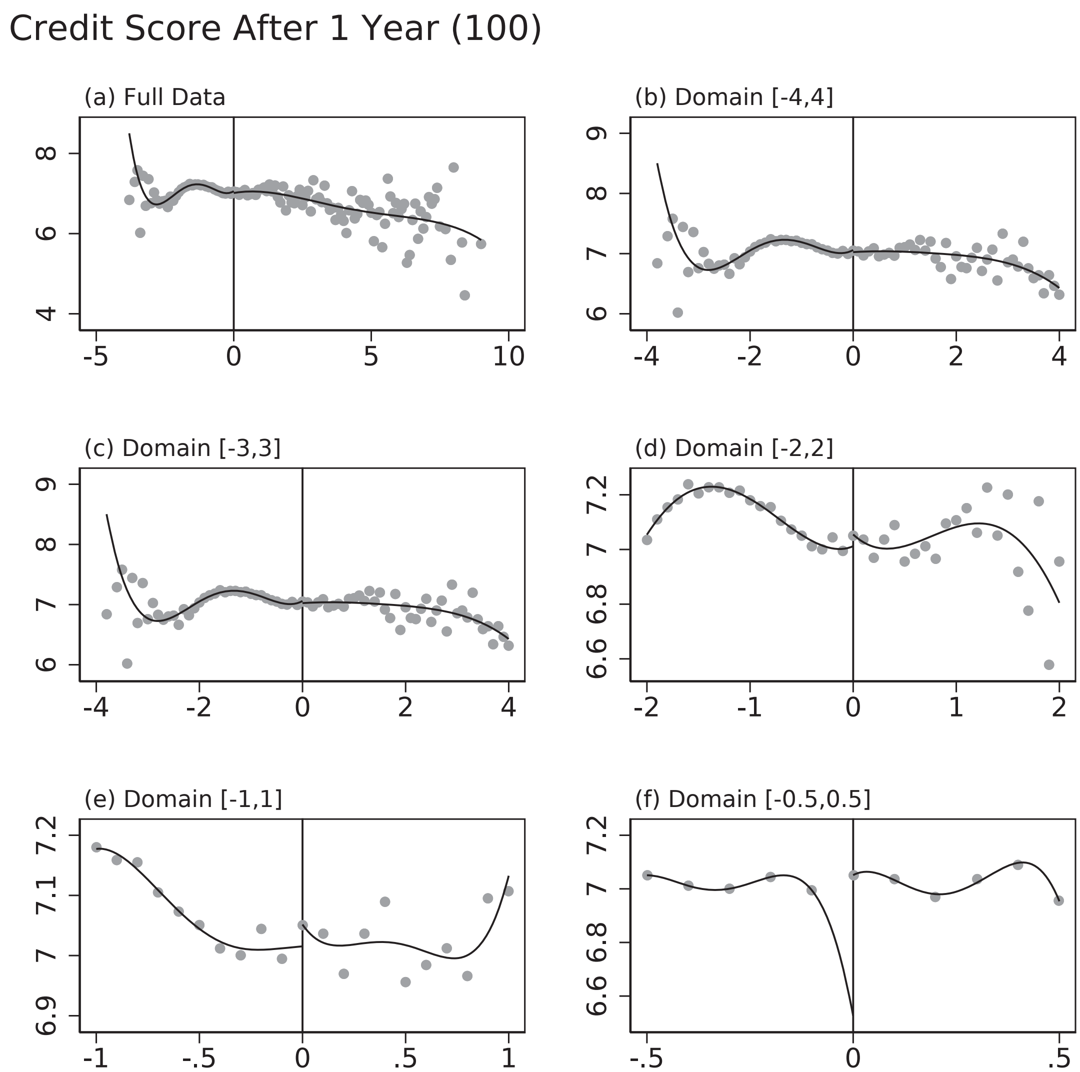

Figure C.1: Bandwidth sensitivity for credit score after 1 year

Note: This figure zooms in on Figure 2.19 between 6.5% A1C ± 2%. It selects a fourth-degree polynomial in all panels, and varies the bandwidth, suggesting that results are sensitive to bandwidth and to order of polynomial used to fit the data.

Appendix D: Robustness of presented RDD models to alternative bandwidths and empirical specifications, Chapter 2

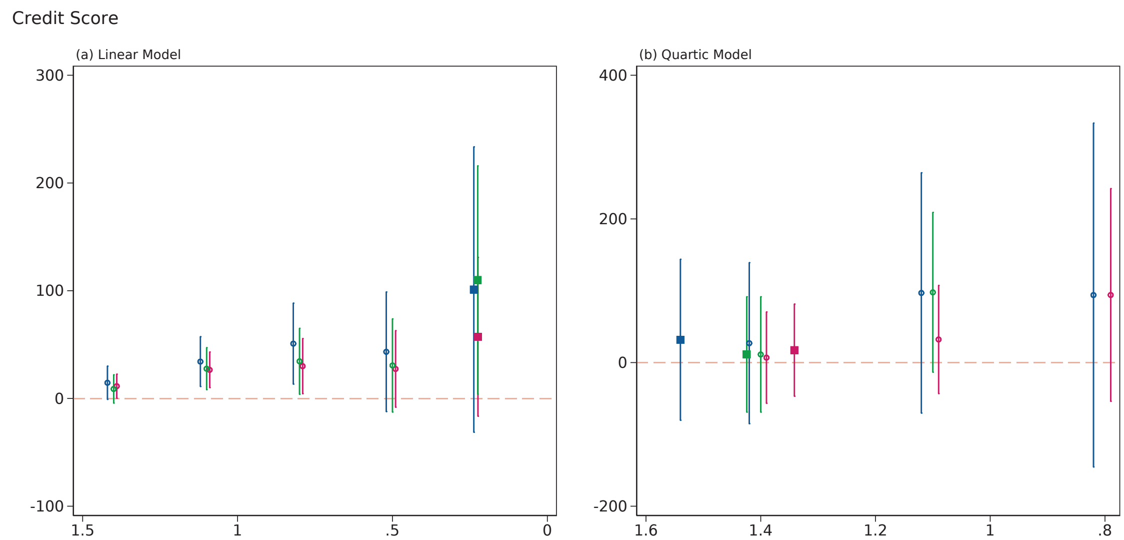

Figure D.1: Robustness of credit score RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

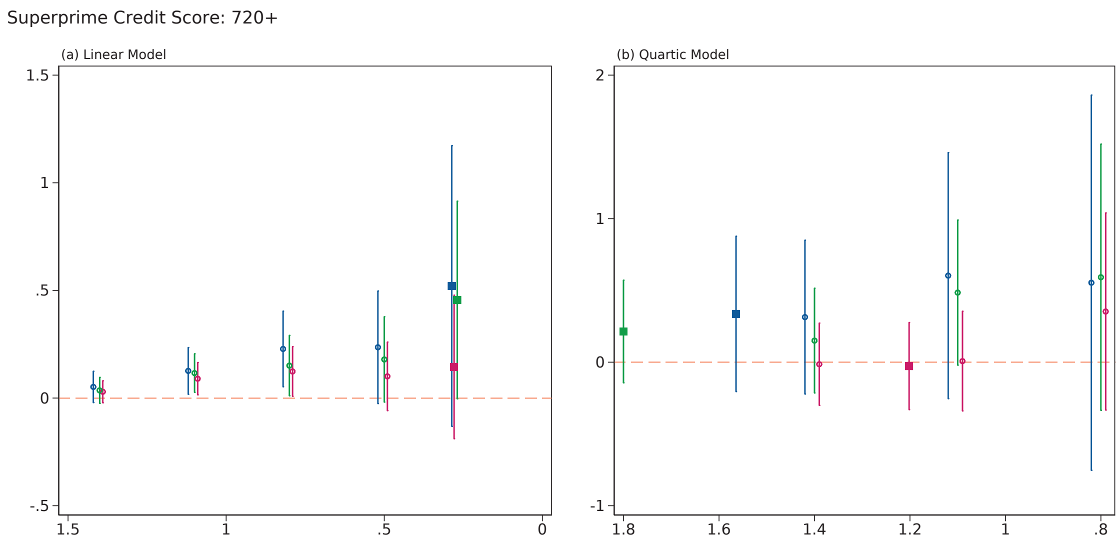

Figure D.2: Robustness of superprime credit score RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

Any 60+ Day Delinquency

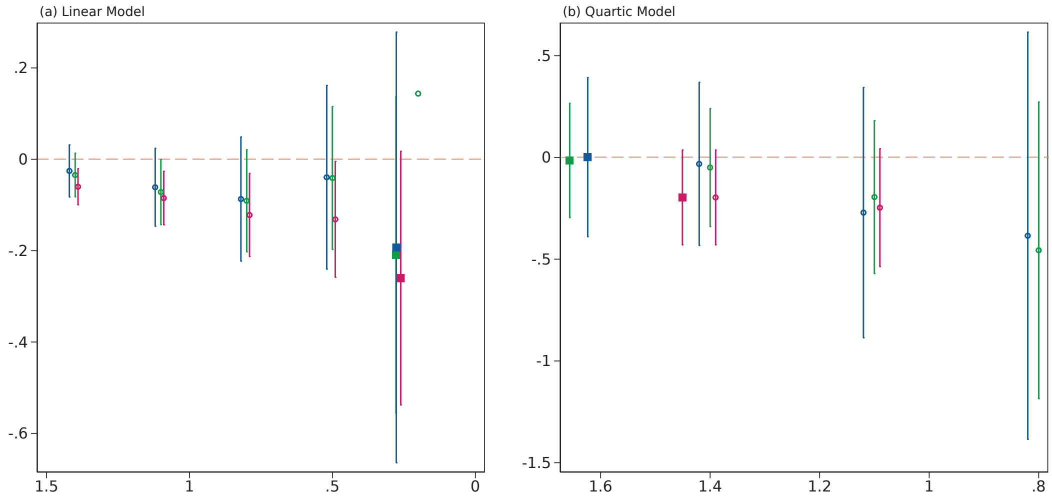

Figure D.3: Robustness of delinquency RD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

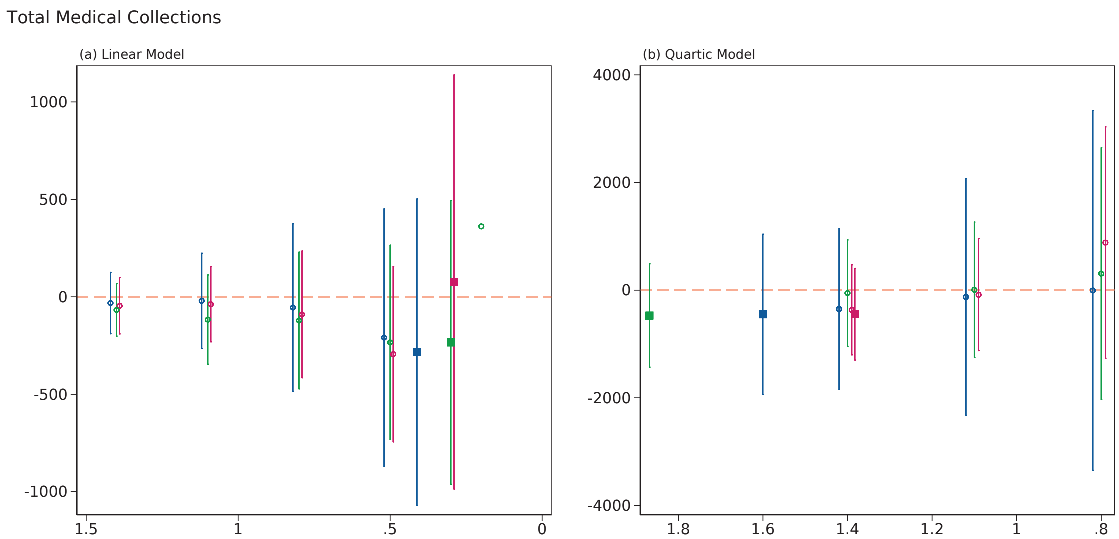

Figure D.4: Robustness of medical collections RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

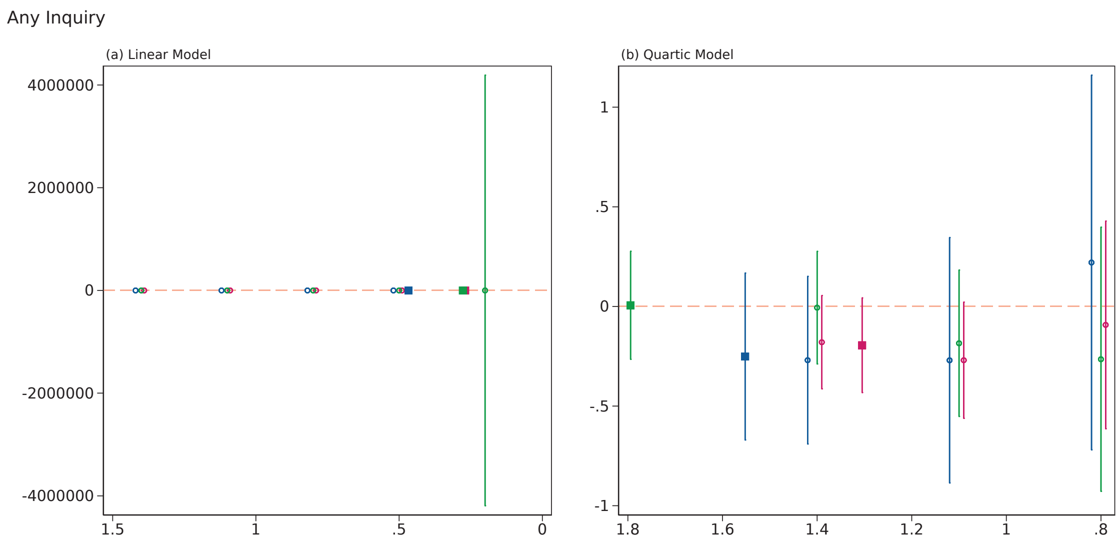

Figure D.5: Robustness of new credit inquiry RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

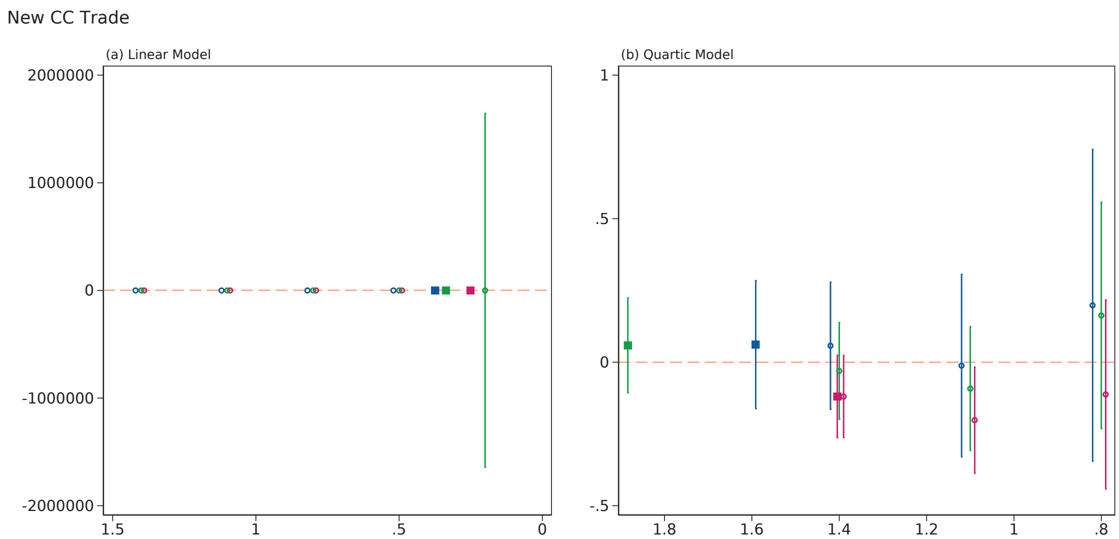

Figure D.6: Robustness of new credit card RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

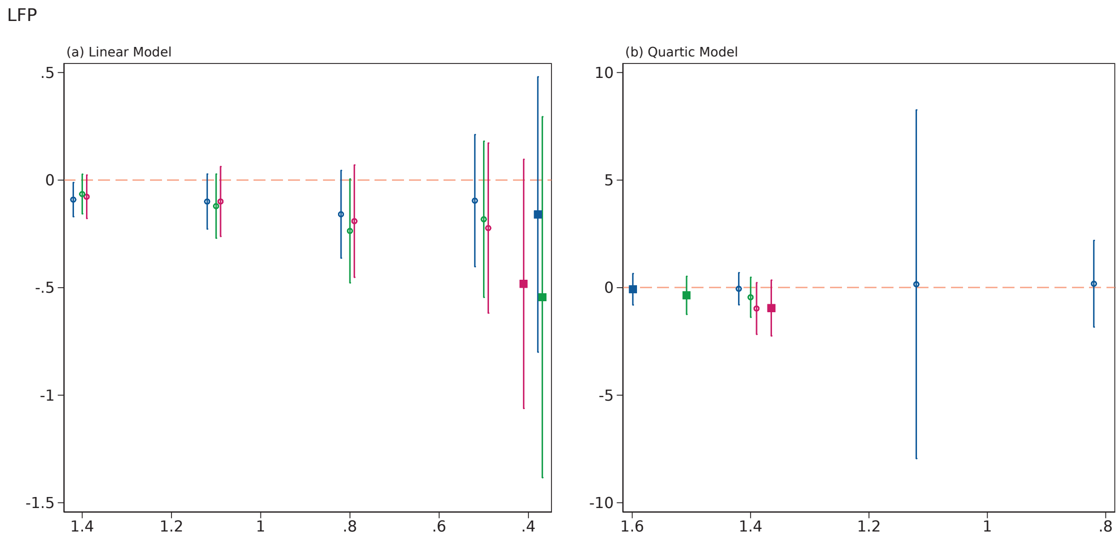

Figure D.7: Robustness of employment rate RD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

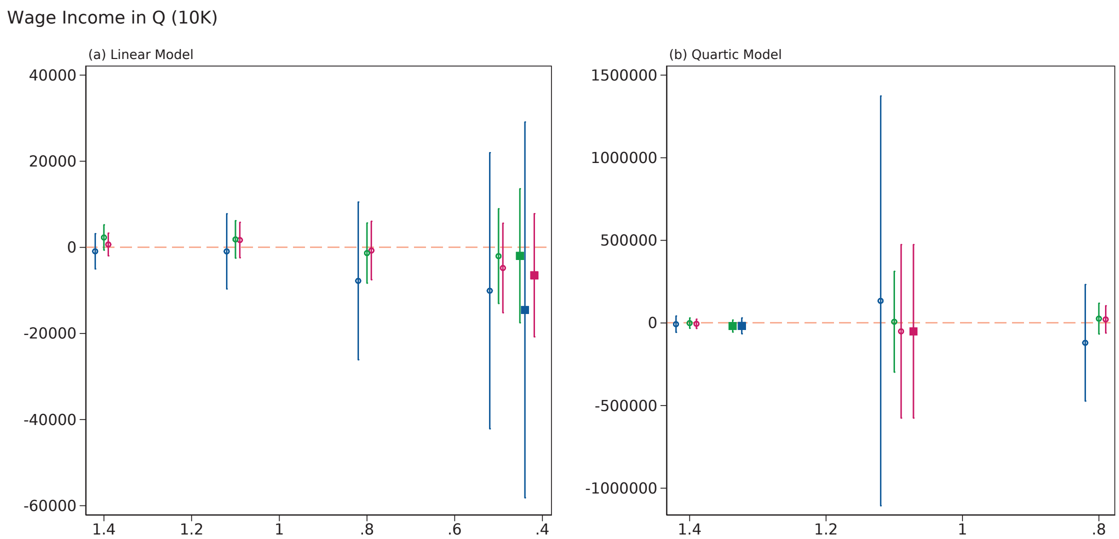

Figure D.8: Robustness of wages RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.

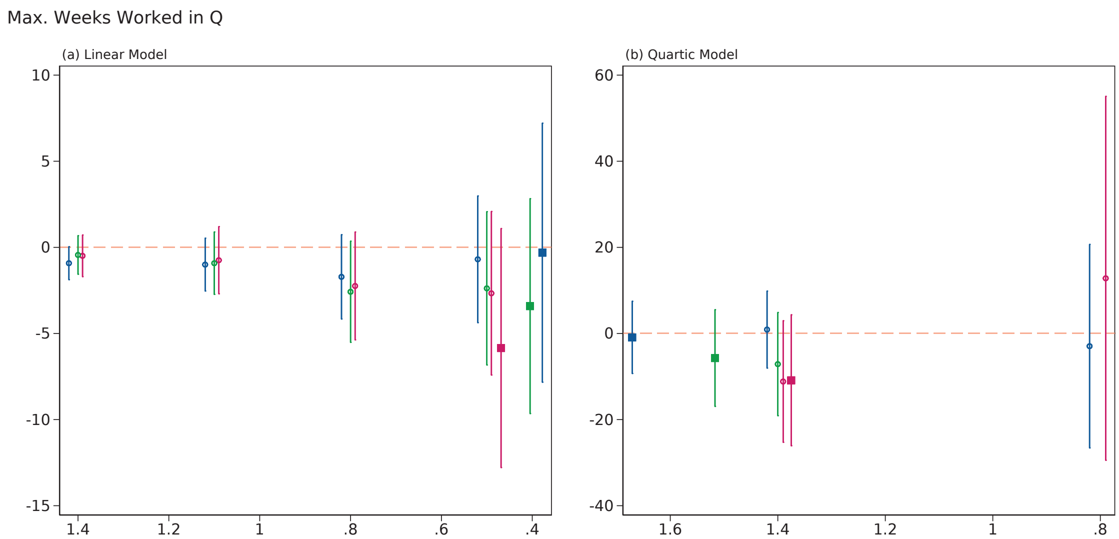

Figure D.9: Robustness of weeks worked RDD to alternative bandwidths and functional forms

Note: This figure presents robustness to the result for a linear and quartic specification, and for alternative bandwidths, where estimation is possible. All models use the RDRobust package in STATA. The cyclamen lines are for the outcome after 1 year, the green line is for the outcome after 2 years, and the blue line is for the outcome after 3 years. The square symbol are the results with the empirically-selected bandwidth for the given model specification.