Causal Analysis of Debt-to-Income Ratio on Loan Default Risk: An Instrumental Variables Approach

Wenxuan Zhu, Rajvi Jasani, Yiqiao Zhu
BIOSTAT 681 Project Report — December 2025

1 Introduction

This section provides an overview of the research problem, explains why endogeneity prevents a simple regression-based analysis of the effect of debt-to-income ratio (DTI) on loan default, and reviews the relevant econometric and credit-market literature motivating the use of instrumental variables (IV). We first present the institutional and empirical background, and then summarize prior research that informs our identification strategy.

1.1 Background

Understanding the causal relationship between a borrower's debt-to-income ratio (DTI) and the probability of loan default is an important problem in credit risk modeling and mortgage underwriting. DTI is a central measure of borrower leverage, but it is also shaped by screening practices, lending constraints, borrower behavior, and unobserved characteristics such as financial discipline or risk preferences. These unobserved factors may simultaneously influence DTI and default likelihood, making DTI an endogenous variable. As a result, ordinary least squares (OLS) estimates provide correlations rather than causal effects, and are biased when unobservable borrower traits drive both repayment capacity and debt levels.

To address this endogeneity, we begin by considering an instrumental-variables (IV) strategy that leverages institutional features of the mortgage market to generate quasi-exogenous variation in DTI. Two potential instruments are present in the data: (1) conforming loan limit status, which constrains borrowing capacity through regulatory thresholds, and (2) pre-approval status, which reflects upstream lender screening and influences the loan terms a borrower is offered. These institutional rules have been used in prior research as sources of exogenous variation because they shift borrowing capacity independently of borrower default risk. Although we initially considered both instruments in parallel, their validity is formally assessed in a dedicated section below, and our final empirical specification retains only the instrument that satisfies the required assumptions.

1.2 Literature Review

Our approach builds on the econometric literature on endogeneity and instrumental-variables methods. Foundational work by Angrist and Imbens (1994) and Angrist and Pischke (2009) establishes that IV estimators identify the Local Average Treatment Effect (LATE) for borrowers whose treatment—here, DTI—is shifted by the instrument. These contributions emphasize the importance of the exclusion restriction, instrument relevance, and monotonicity, and highlight the bias that arises when treatment variables correlate with unobserved heterogeneity. Standard references such as Wooldridge (2010) further reinforce the need for causal inference tools in observational credit data.

In credit-market applications, borrower leverage and underwriting thresholds play a central role in determining credit access and loan performance. Research on mortgage underwriting (Avery et al., 1996; Gerardi et al., 2008; Agarwal et al., 2017) demonstrates that DTI caps, screening rules, and loan-to-value constraints reflect both institutional lending policies and individual borrower characteristics, reinforcing concerns that DTI is endogenous. Studies of lender screening and borrower selection (Keys et al., 2010; Campbell et al., 2011; Bhutta and Keys, 2016) document that pre-approval processes and screening rules meaningfully shape borrower outcomes, motivating the use of pre-approval as a potential instrument.

Regulatory loan limits also generate quasi-exogenous variation in borrowing capacity. Conforming loan thresholds affect permissible loan size and underwriting criteria (Bhutta et al., 2020) and have been employed in prior quasi-experimental studies of credit access and mortgage performance (Adelino et al., 2016; Di Maggio et al., 2017). Together, this literature supports the use of institutional lending rules as potential instruments for identifying the causal effect of DTI on default. In the analysis that follows, we evaluate instrument validity empirically and retain only the instrument that satisfies the necessary assumptions.

2 Methodology

This section describes our analytical approach for estimating the causal effect of debt-to-income ratio (DTI) on loan default probability. We address two key methodological challenges: (1) substantial missing data in DTI with a non-random pattern, and (2) potential endogeneity of DTI due to unobserved borrower characteristics. Our methodology combines multiple imputation by chained equations (MICE) to handle missing data with two-stage least squares (2SLS) instrumental variables estimation to address endogeneity.

2.1 Data and Sample

Our analysis uses a comprehensive bank loan dataset from 2019, originally comprising 148,670 loan records and 34 variables capturing applicant-level and loan-level attributes. The outcome variable is a binary indicator of default status, where $Y_i = 1$ denotes default and $Y_i = 0$ denotes successful repayment. The treatment variable is the debt-to-income ratio (DTI), a continuous measure representing the proportion of income allocated to debt service. The overall default rate in the sample is 24.6%.

2.2 Missing Data Analysis

2.2.1 Identifying the Missing Data Mechanism

A critical challenge in our analysis is the substantial missingness in the DTI variable. We observed a pronounced differential pattern: only 7.0% of non-defaulting loans had missing DTI values, compared to 44.5% of defaulting loans. This stark asymmetry suggests a Missing Not At Random (MNAR) mechanism, where missingness is directly related to the outcome.

We formally tested this pattern using a chi-square test of independence between default status and DTI missingness:

$$H_0 : \text{DTI missingness is independent of default status} \tag{1}$$

The test yielded $\chi^2 = 28{,}628$ with $p < 0.001$, providing strong evidence that DTI missingness is systematically related to default status. This MNAR pattern has important implications: complete case analysis would systematically exclude the highest-risk borrowers, biasing our estimates toward null.

2.2.2 Multiple Imputation by Chained Equations (MICE)

To address the MNAR pattern while retaining maximum sample size and preserving variability, we employed Multiple Imputation by Chained Equations (MICE). Unlike single imputation methods, MICE generates multiple plausible datasets, each with different imputed values, allowing proper propagation of imputation uncertainty to final estimates.

The MICE algorithm operates iteratively. For each variable X_j with missing values, we specify a conditional imputation model:

$$X_j^{(t+1)} \sim f(X_j \mid X_{-j}^{(t)}, \theta_j) \tag{2}$$

where $X_{-j}^{(t)}$ represents all other variables at iteration t, and θ_j are model parameters. The algorithm cycles through all variables with missing data until convergence, then draws from the posterior predictive distribution to generate imputed values.

Imputation Model Specification. We used predictive mean matching (PMM) for all continuous variables (DTI, credit score, income, property value, interest rate, and loan term). PMM is advantageous because it preserves the original data distribution and ensures imputed values fall within plausible ranges by matching predicted values to observed donor values.

A key feature of our imputation strategy is the inclusion of the outcome variable (default status) as a predictor in all imputation models. This approach is essential when missingness depends on the outcome, as in our MNAR setting. By conditioning on default status, the imputation model can capture the relationship between missingness and outcome, generating higher imputed DTI values for defaulting loans consistent with the observed pattern.

Implementation Details. We generated $m = 5$ imputed datasets with 10 iterations per imputation (maxit = 10). Convergence was assessed visually through trace plots of imputed values across iterations. The imputed datasets showed stable estimates after approximately 5 iterations, indicating adequate convergence. Validation checks confirmed that imputed DTI distributions were similar to observed distributions while appropriately reflecting the MNAR pattern: mean imputed DTI for defaults (39.71) exceeded that for non-defaults (37.69).

2.3 Instrumental Variables Framework

2.3.1 The Endogeneity Problem

Standard ordinary least squares (OLS) regression of default on DTI produces biased estimates because DTI is likely correlated with unobserved determinants of default:

$$Y_i = \beta_0 + \beta_1 \mathrm{DTI}_i + X_i'\gamma + \varepsilon_i \tag{3}$$

where $E[\mathrm{DTI}_i \cdot \varepsilon_i] \neq 0$. Borrowers with higher unobserved risk profiles (e.g., financial instability, poor money management) may simultaneously have higher DTIs and higher default probabilities, inducing positive correlation between DTI and the error term. This endogeneity causes OLS to conflate the causal effect of DTI with selection effects.

2.3.2 Instrumental Variables Approach

To identify the causal effect, we employ two-stage least squares (2SLS) with an instrumental variable Z that satisfies two conditions:

1. **Relevance:** $\mathrm{Cov}(Z, \mathrm{DTI}) \neq 0$ — the instrument must predict DTI
2. **Exclusion Restriction:** $\mathrm{Cov}(Z, \varepsilon) = 0$ — the instrument affects default only through DTI

The 2SLS procedure estimates the causal effect in two stages:

First Stage.

$$\mathrm{DTI}_i = \pi_0 + \pi_1 Z_i + X_i'\delta + \nu_i \tag{4}$$

This regression isolates variation in DTI driven by the instrument Z, purged of correlation with unobserved confounders.

Second Stage.

$$Y_i = \beta_0 + \beta_1 \widehat{\text{DTI}}_i + X_i'\gamma + \eta_i \tag{5}$$

where $\widehat{\text{DTI}}_i$ is the fitted value from the first stage. Since $\widehat{\text{DTI}}_i$ contains only exogenous variation in DTI, $\hat{\beta}_1$ consistently estimates the causal effect.

2.3.3 Instrument Selection and Validation

Our initial specification included two candidate instruments: (1) `loan_limit`, indicating conforming versus non-conforming loan status, and (2) `approv_in_adv`, indicating pre-approval status.

Economic Rationale. Conforming loans adhere to regulatory criteria (e.g., Fannie Mae/Freddie Mac standards) that constrain permissible DTI thresholds, providing a source of variation in DTI that is plausibly unrelated to borrower-specific default risk. Pre-approval status reflects institutional screening that influences acceptable DTI levels through different underwriting standards. Both instruments should affect default probability only through their effect on DTI, conditional on observable borrower characteristics.

Testing Instrument Validity. With two instruments and one endogenous variable, our initial model was overidentified, allowing us to test instrument validity using the Sargan test of overidentifying restrictions. The null hypothesis is that all instruments are valid (uncorrelated with the structural error). In our analysis, the Sargan test rejected instrument validity at the 5% level ($p \approx 0.044$), indicating that at least one instrument violates the exclusion restriction.

Individual instrument tests revealed that `approv_in_adv` was the stronger instrument (partial $F = 27.1$) compared to `loan_limit` (partial $F = 14.9$). Further investigation showed that both instruments significantly predicted loan-to-value (LTV) ratio, suggesting LTV may be a post-treatment variable that should not be included as a control. After removing LTV from the control set, the joint first-stage F-statistic increased from 8.34 to 18.88.

Final Instrument Selection. Given the Sargan test rejection, we proceeded with a single-instrument specification using only `approv_in_adv`. This yields an exactly identified model where the Sargan test is not applicable but instrument validity relies on economic reasoning. Pre-approval status creates variation in DTI through institutional screening policies that are plausibly exogenous to individual borrower default risk, conditional on our control variables.

2.4 Estimation Strategy

2.4.1 Model Specification

Our final specification estimates the following 2SLS model on each imputed dataset:

First Stage:

$$\text{DTI}_i = \pi_0 + \pi_1 \cdot \mathbf{1}[\text{pre-approval}_i] + X_i'\delta + \nu_i \tag{6}$$

Second Stage:

$$\text{Default}_i = \beta_0 + \beta_1 \widehat{\text{DTI}}_i + X_i'\gamma + \eta_i \tag{7}$$

The control vector X_i includes credit score, loan amount, income, interest rate, property value, loan term, gender, region, loan type, credit worthiness, occupancy type, and age group. We exclude LTV from the control set based on evidence that it is affected by the instrument and thus represents a post-treatment variable.

2.4.2 Pooling Across Imputations

We estimate the 2SLS model separately on each of the $m = 5$ imputed datasets and combine results using Rubin's rules. Let $\hat{\beta}_1^{(j)}$ and $\hat{V}^{(j)}$ denote the DTI coefficient estimate and its variance from imputed dataset j.

Pooled Point Estimate:

$$\bar{\beta}_1 = \frac{1}{m}\sum_{j=1}^{m}\hat{\beta}_1^{(j)} \tag{8}$$

Pooled Variance: The total variance combines within-imputation variance ($\bar{W}$) and between-imputation variance (B):

$$T = \bar{W} + \left(1 + \frac{1}{m}\right)B \tag{9}$$

where

$$\bar{W} = \frac{1}{m}\sum_{j=1}^{m}\hat{V}^{(j)}, \quad B = \frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{\beta}_1^{(j)} - \bar{\beta}_1\right)^2 \tag{10}$$

The factor $(1 + 1/m)$ accounts for finite-m correction. Inference proceeds using a t-distribution with degrees of freedom approximated by:

$$\nu = (m-1)\left(1 + \frac{\bar{W}}{(1+1/m)B}\right)^2 \tag{11}$$

2.5 Diagnostic Tests

2.5.1 Instrument Strength

We assess first-stage instrument relevance using the F-statistic from testing $H_0 : \pi_1 = 0$ in the first-stage regression. Following the rule of thumb from Staiger and Stock (1997), $F > 10$ indicates a strong instrument, while F between 4 and 10 suggests moderate strength with potential weak-instrument concerns. We report the mean, minimum, and maximum F-statistics across imputed datasets.

2.5.2 Endogeneity Test

The Wu-Hausman test assesses whether OLS and IV estimates are statistically different, with the null hypothesis that DTI is exogenous. Rejection of the null supports the IV approach. We report the proportion of imputed datasets where the Wu-Hausman test is significant at $\alpha = 0.05$.

2.5.3 Covariate Balance

Although not required for IV validity (which relies on conditional exogeneity), we examine covariate balance by testing whether the instrument predicts pre-treatment covariates. Significant imbalance motivates inclusion of these covariates as controls to support conditional instrument validity.

2.6 Robustness Checks

We conduct several sensitivity analyses to assess the robustness of our findings:

1. **Complete Case Analysis:** We estimate the model using only observations with complete DTI data to assess sensitivity to imputation assumptions.
2. **OLS vs. IV Comparison:** We compare pooled OLS and IV estimates to quantify the magnitude and direction of endogeneity bias.
3. **Imputation Validation:** We compare distributions of observed and imputed DTI values, overall and stratified by default status, to verify that imputed values are plausible.

2.7 Interpretation: Local Average Treatment Effect

Under the IV framework, our estimate identifies the Local Average Treatment Effect (LATE) rather than the Average Treatment Effect (ATE). Since our treatment (DTI) is continuous, our IV estimate captures a weighted average of causal effects across the distribution of DTI shifts induced by the instrument. The LATE represents the causal effect of DTI on default for "compliers"—borrowers whose DTI levels are affected by pre-approval status. This complier population may differ from the overall population, limiting external validity. However, the LATE remains a valid causal effect for a well-defined subpopulation and is policy-relevant for understanding how institutional screening policies affect default risk through their impact on debt burden.

3 Results

3.1 Missing Data Patterns

Table 1 presents the distribution of missingness across key variables, stratified by loan default status. The debt-to-income ratio (DTI) exhibits a pronounced differential pattern: only 7.0% of non-defaulting loans have missing DTI values, compared to 44.5% of defaulting loans. This stark asymmetry provides strong evidence of a Missing Not At Random (MNAR) mechanism.

Table 1: Missing Data Patterns by Default Status

Missing Status	N	DTI	DTI %	Approv	Credit	Income
Non-defaults	112,031	7,811	7.0%	0	0	7,911
Defaults	36,639	16,310	44.5%	0	0	1,239

We formally tested the independence of DTI missingness and default status using a chi-square test. The test yielded $\chi^2 = 28{,}628$ with $p < 0.001$, providing overwhelming evidence that missingness is systematically related to the outcome. This MNAR pattern justifies our use of Multiple Imputation by Chained Equations (MICE) with the outcome variable included as a predictor in the imputation model.

Our instrumental variable (`approv_in_adv`) exhibits complete data availability (0% missing) in both groups, ensuring that instrument validity assessments are not compromised by selection. Other finance-related covariates also display differential missingness, supporting a multivariate imputation strategy. After initial data preparation and excluding observations with missing outcome or instrument values, the analytical sample comprised 147,762 loans, with DTI observed for 123,769 (83.8%) and missing for 23,993 (16.2%).

Table 2 summarizes the extent of missingness among all variables requiring imputation.

Table 2: Variables Requiring Imputation

Variable	N Missing	% Missing
Interest Rate	36,198	24.5%
DTI	23,993	16.2%
Property Value	15,001	10.2%

3.2 Imputation Validation

Following the MICE procedure described in Section 3.3, we generated five imputed datasets with 10 iterations per imputation. Convergence diagnostics, as seen in Fig. 1, confirmed stable estimates after approximately 5 iterations.

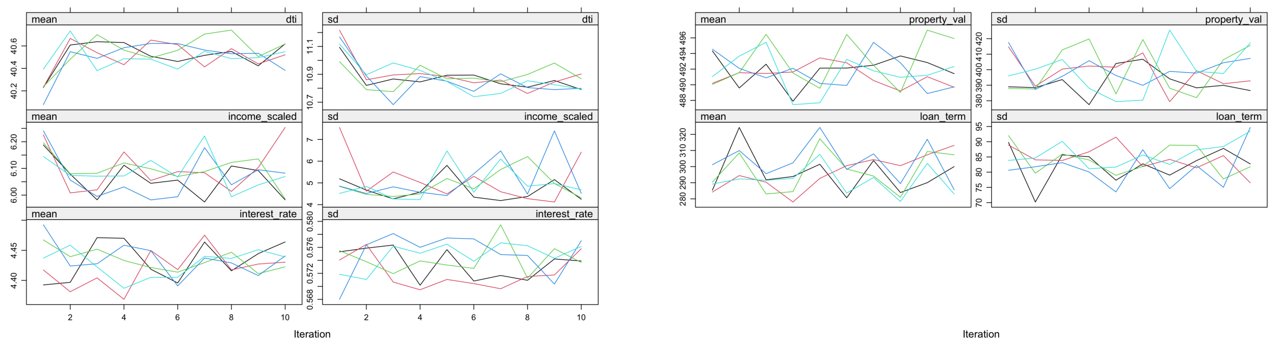

Fig. 1: Convergence Plots for Imputed Variables across Iterations

Fig. 2 compares the distributions of observed and imputed DTI values across default-status groups, providing evidence on the quality of the imputations. Table 3 presents mean DTI values, revealing that the imputation model successfully captured the relationship between DTI and default status. Imputed DTI values for defaulted loans (mean = 39.71) appropriately exceed those for non-defaults (mean = 37.69), consistent with the MNAR mechanism where high-DTI defaults are disproportionately missing.

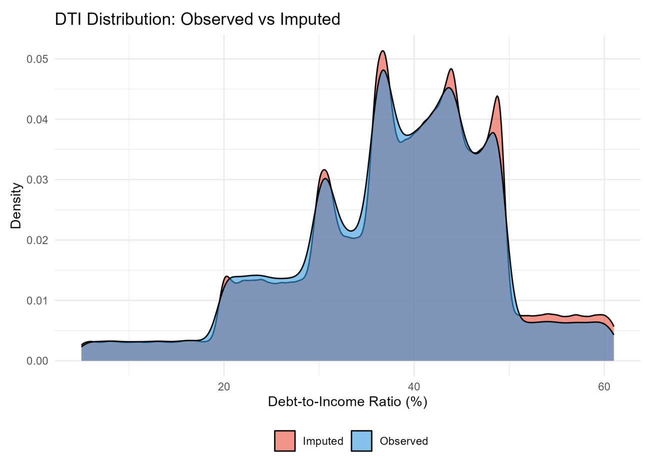

Fig. 2: Validation for DTI Imputation

Table 3: Mean Imputed DTI by Default Status

Status	Mean DTI
Non-defaults	37.69
Defaults	39.71

This pattern confirms that the outcome-conditional imputation appropriately generates higher DTI values for missing defaults, thereby addressing the MNAR bias that would arise from complete-case analysis.

3.3 Complete-Case Benchmark Results

Before proceeding with the primary MICE-based analysis, we estimated both OLS and IV models on the complete-case sample ($n = 123{,}659$ with observed DTI) to establish baseline comparisons.

3.3.1 First-Stage Diagnostics

The first-stage regression shows that pre-approval status is a strong predictor of DTI. The first-stage F-statistic is 20.72 ($p < 0.001$), substantially exceeding the conventional threshold of 10 for strong instruments. This confirms instrument relevance in the complete-case sample.

3.3.2 Endogeneity Test

The Wu-Hausman test strongly rejects the null hypothesis of exogeneity ($p < 0.001$), providing formal evidence that DTI is endogenous. This confirms that OLS estimates are inconsistent and that instrumental variables estimation is necessary for valid causal inference.

3.3.3 Complete-Case Estimates

The complete-case OLS estimate indicates a positive association: $\hat{\beta}_{\text{OLS}} = 0.001543$ (SE $= 0.000106$, $p < 0.001$). However, given the rejection of exogeneity, this coefficient cannot support causal interpretation.

The complete-case IV estimate reveals a significant negative effect: $\hat{\beta}_{\text{IV}} = -0.1186$ (SE $= 0.0276$, $p < 0.001$, 95% CI: $[-0.173, -0.064]$). This indicates that the causal effect of DTI on default probability is negative within the complier population, even in the restricted complete-case sample.

3.4 Primary Analysis: Pooled MICE Results

We estimated both OLS and 2SLS models on each of the five multiply imputed datasets and combined results using Rubin's rules. This section presents our primary findings.

3.4.1 Instrument Diagnostics Across Imputations

The instrument remains consistently strong across all five imputations. First-stage F-statistics range from 15.25 to 21.91, with a mean of 18.95, well above the threshold of 10 for strong instruments. The Wu-Hausman test rejects exogeneity in all five imputations ($p < 0.001$ in each), confirming that IV estimation is necessary regardless of the missing data approach.

These diagnostics validate both the instrument strength and the need for IV estimation in the imputed datasets.

3.4.2 Pooled IV Results

Table 4 presents the pooled 2SLS estimates for key variables. The pooled IV estimate for DTI is $\hat{\beta}_{\text{IV}} = -0.14298$ (SE = 0.03807, $p = 0.0002$, 95% CI: $[-0.218, -0.068]$). This indicates that a one-percentage-point increase in DTI *causally reduces* the probability of default by approximately 14.3 percentage points for the complier population.

Table 4: Pooled 2SLS Results using MICE

Variable	**Estimate**	**SE**	***t*-stat**	***p*-value**	**CI Lower**	**CI Upper**	**Sig.**
DTI	−0.14298	0.03807	−3.756	0.0002	−0.218	−0.068	***
Credit Score	0.00001	0.00004	0.274	0.784	−0.00006	0.00008	
Loan Amount	0.00177	0.00046	3.892	0.0001	0.00088	0.00267	***
Income (scaled)	−0.07762	0.01997	−3.887	0.0001	−0.117	−0.038	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

The negative causal effect stands in sharp contrast to conventional credit risk intuition. However, this finding has a coherent interpretation within the Local Average Treatment Effect (LATE) framework. Pre-approval status identifies borrowers who have passed stricter institutional screening, signaling superior creditworthiness. While pre-approval may mechanically increase approved loan amounts (and thus DTI), these borrowers' underlying default risk remains lower. Consequently, the instrument-induced variation in DTI is negatively correlated with default probability, yielding a negative LATE estimate for compliers.

3.4.3 Pooled OLS Results

For comparison, Table 5 presents the pooled OLS estimates. The pooled OLS estimate is $\hat{\beta}_{\text{OLS}} = 0.002177$ (SE = 0.000183, $p < 0.001$), indicating a positive association. However, because the Wu–Hausman test rejects exogeneity in all imputations, this coefficient is inconsistent and confounded by selection bias.

Table 5: Pooled OLS Results using MICE

Variable	**Estimate**	**SE**	***t*-stat**	***p*-value**	**CI Lower**	**CI Upper**	**Sig.**
DTI	0.002177	0.000183	11.925	<0.001	0.00182	0.00254	***
Credit Score	0.00002	0.00001	1.607	0.108	−0.000003	0.00003	
Loan Amount	0.00003	0.000007	3.678	0.0003	0.00001	0.00004	***
Income (scaled)	−0.00191	0.000226	−8.418	<0.001	−0.00235	−0.00146	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

The stark difference between OLS (+0.002177) and IV (−0.14298) estimates reveals severe endogeneity bias. OLS conflates the causal effect with selection effects, producing substantially attenuated (and sign-reversed) estimates.

3.5 Comparison Across Estimation Strategies

Fig. 3 presents DTI effect estimates across four specifications: complete-case OLS, complete-case IV, MICE OLS, and MICE IV. This comparison allows us to assess both the impact of endogeneity correction (OLS vs. IV) and the impact of missing data strategy (complete-case vs. MICE). Detailed summary of estimates is shown in Table 6

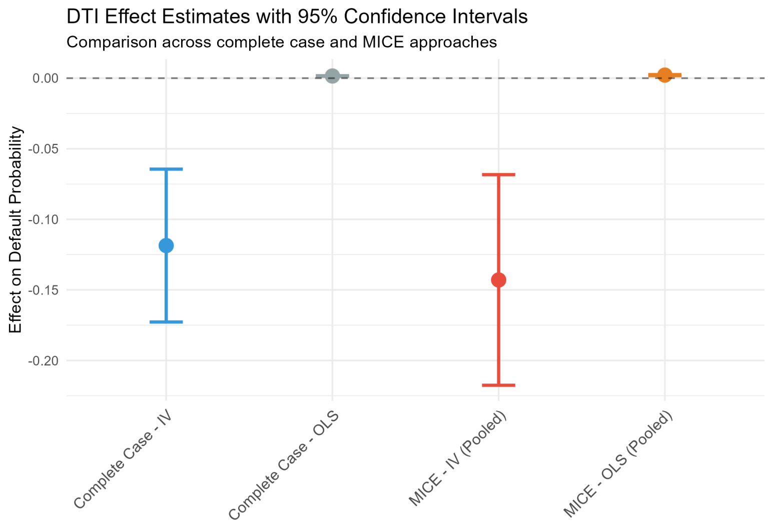

Fig. 3: Comparison of Different Analysis Strategies

Table 6: DTI Effect Estimates across Methods

Approach	N	Defaults	Coef.	SE	CI Lower	CI Upper	Sig.
Complete Case – OLS	123,659	20,177	0.001543	0.000106	0.00134	0.00175	***
Complete Case – IV	123,659	20,177	−0.1186	0.0276	−0.173	−0.064	***
MICE – OLS (Pooled)	147,762	36,398	0.002177	0.000183	0.00182	0.00254	***
MICE – IV (Pooled)	147,762	36,398	−0.14298	0.03807	−0.218	−0.068	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Several patterns emerge from this comparison:

OLS vs. IV (Within Method). Within both complete-case and MICE approaches, IV estimates are substantially larger in magnitude and opposite in sign compared to OLS estimates. This pattern confirms severe endogeneity bias: OLS conflates causal effects with selection effects.

Complete-Case vs. MICE (Within Estimator). The IV estimates are qualitatively consistent across missing data strategies: both are negative and statistically significant. However, the MICE-IV estimate (-0.14298) has a slightly larger magnitude than the complete-case IV estimate (-0.1186), with wider confidence intervals reflecting additional imputation uncertainty. Notably, the MICE approach utilizes 147,762 observations (including 36,398 defaults) compared to only 123,659 observations (20,177 defaults) in the complete-case analysis, a gain of 24,103 observations and 16,221 defaults.

Interpretation. The qualitative conclusion, that DTI has a negative causal effect on default for the complier population, is robust across missing data strategies. However, complete-case analysis systematically excludes high-risk defaults (44.5% of defaults have missing DTI), potentially attenuating the estimated effect. The MICE-IV specification represents our preferred estimate because it: (1)

uses the full sample, maximizing statistical power; (2) addresses MNAR through outcome-conditional imputation; and (3) properly propagates imputation uncertainty.

3.6 Covariate Balance Tests

To evaluate the plausibility of the conditional exclusion restriction, we tested whether pre-approval status predicts pre-treatment covariates. Table 7 presents results from regressing each covariate on the instrument.

Table 7: Covariate Balance Tests: Pre-Approval Status vs. Pre-Treatment Covariates

Instrument	**Covariate**	**Estimate**	**SE**	***p*-value**	**Balanced**
approv_adv	Credit Score	-0.378	0.829	0.649	
approv_adv	Loan Amount	-15.222	1.316	<0.001	×
approv_adv	Income	-0.200	0.046	<0.001	×

Pre-approval status is not significantly associated with credit score ($p = 0.649$) but exhibits significant associations with loan amount and income ($p < 0.001$ for both). These imbalances do *not* invalidate the IV strategy, as instrumental variables estimation requires only *conditional* independence. Both loan amount and income are explicitly included as controls in our regressions, absorbing their association with the instrument. The exclusion restriction requires only that, conditional on these controls, pre-approval affects default solely through DTI, a restriction supported by economic reasoning regarding institutional screening policies.

3.7 Summary of Main Findings

Our analysis yields the following primary results:

1. **MNAR Pattern:** DTI exhibits pronounced differential missingness (7.0% among non-defaults vs. 44.5% among defaults; $\chi^2 = 28{,}628$, $p < 0.001$), necessitating outcome-conditional imputation.

2. **Strong Instrument:** Pre-approval status consistently predicts DTI across all imputations, with first-stage F-statistics ranging from 15.25 to 21.91 (mean = 18.95).

3. **Confirmed Endogeneity:** The Wu–Hausman test rejects exogeneity in all specifications ($p < 0.001$), validating the necessity of IV estimation.

4. **Primary Causal Estimate:** The pooled MICE-IV estimate is $\hat{\beta}_{\text{IV}} = -0.14298$ (SE = 0.03807, $p = 0.0002$), indicating that a one-percentage-point increase in DTI causally reduces default probability by 14.3 percentage points for compliers.

5. **Robustness:** The negative causal effect is qualitatively robust across missing data strategies (complete-case: -0.1186; MICE: -0.14298), though complete-case analysis yields attenuated estimates due to systematic exclusion of high-risk defaults.

6. **Local Interpretation:** The estimate identifies a LATE for borrowers whose DTI is affected by pre-approval screening, creditworthy borrowers who receive higher loan approvals but maintain lower underlying default risk.

These findings provide causal evidence that, for the complier population, higher DTI ratios induced by pre-approval screening are associated with *lower* default probabilities, reflecting the selection of creditworthy borrowers through institutional screening processes.

4 Discussion

4.1 Summary and Interpretation of Findings

Our IV analysis yields a striking result: a one percentage point increase in DTI *reduces* default probability by approximately 14.3 percentage points ($\hat{\beta}_{IV} = -0.143$, $p = 0.0002$), sharply contrasting with the small positive OLS estimate ($\hat{\beta}_{OLS} = 0.002$). The Wu-Hausman test rejects exogeneity across all five imputed datasets, confirming that IV estimation is necessary for causal inference.

4.1.1 Reconciling the Sign Reversal

The sign reversal from OLS to IV warrants careful interpretation. The positive OLS association reflects both causal effects and selection: borrowers with unobserved risk factors (e.g., financial instability) may simultaneously accumulate higher debt and face elevated default risk.

The IV estimate isolates variation in DTI induced by pre-approval status, which reflects institutional screening rather than borrower-driven debt accumulation. Pre-approved borrowers have passed stricter underwriting, signaling greater creditworthiness. These borrowers may receive larger approved loan amounts, mechanically increasing DTI, while their underlying default risk remains low. This produces a negative causal estimate for the complier population.

4.1.2 Magnitude Considerations

The effect magnitude deserves scrutiny. A 14.3 percentage point reduction per one percentage point DTI increase implies implausibly large effects for borrowers with substantial DTI changes. This suggests several possibilities: (1) the CACE applies to a narrow complier population with low baseline default risk; (2) pre-approval selects borrowers with exceptionally favorable unobserved characteristics; (3) finite-sample bias despite adequate first-stage F-statistics (mean $F = 18.95$); or (4) functional form limitations of the linear probability model.

We interpret our results as strong evidence that the positive OLS association is driven by selection bias, and that the causal effect for compliers is likely negative or null. However, the precise magnitude should be interpreted with caution.

4.2 Implications

Our findings challenge the conventional view that higher DTI mechanically increases default risk. While DTI remains a useful underwriting metric, its predictive value may stem primarily from correlation with unobserved borrower characteristics rather than a direct causal pathway. The effectiveness of pre-approval as an instrument highlights that comprehensive screening, incorporating soft information and holistic risk assessment, may be more predictive of default than mechanical DTI cutoffs alone.

For policy, our results suggest that DTI limits in regulatory frameworks (e.g., Qualified Mortgage rules) may function primarily as proxies for broader creditworthiness rather than as direct constraints on unsustainable borrowing.

4.3 Limitations

4.3.1 Instrument Validity

The exclusion restriction is fundamentally untestable. Pre-approval may affect default through channels other than DTI, such as signaling lender confidence or correlating with unobserved loan terms.

With a single instrument, we cannot conduct overidentification tests; the Sargan test rejected joint validity when loan limit was included ($p \approx 0.044$), leaving us with exact identification that hinges entirely on maintained assumptions.

4.3.2 External Validity

Our IV estimate identifies the CACE for borrowers whose DTI is affected by pre-approval, likely marginal borrowers with distinct risk profiles. This complier population may not represent all borrowers, limiting generalizability to broader policy interventions.

4.3.3 Data Limitations

Our single cross-section from 2019 limits temporal generalizability. The dataset lacks detailed borrower financial histories and property appraisal information. The pre-COVID institutional context may limit applicability to current market conditions.

4.4 Future Research

Future work could explore alternative instruments (geographic housing cost variation, policy discontinuities), investigate heterogeneous treatment effects across borrower subgroups (credit score, loan purpose, property type), and employ dynamic models to capture how DTI interacts with post-origination factors. Causal mediation analysis could elucidate mechanisms linking debt burden to default, while comparisons with machine learning approaches could clarify the predictive versus causal role of DTI in credit risk modeling.

4.5 Conclusion

This study provides causal evidence on the DTI-default relationship using instrumental variables combined with multiple imputation. Our key finding, a negative IV estimate contrasting with a positive OLS estimate, demonstrates the importance of addressing endogeneity in credit risk analysis. The naive positive association appears driven by selection rather than a causal mechanism.

While the magnitude of our IV estimate should be interpreted cautiously, the qualitative conclusion is robust: higher DTI does not mechanically cause higher default risk for the complier population. These findings underscore the value of comprehensive underwriting practices that look beyond simple leverage metrics and highlight the need for continued research into the causal determinants of loan default.

References

Adelino, M., Schoar, A., and Severino, F. (2016). Loan originations and credit trends. *Review of Financial Studies*, 29(8):2187–2223.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., and Evanoff, D. (2017). The role of housing equity in mortgage modification decisions. *Journal of Financial Economics*, 125(1):1–17.

Angrist, J. D. and Imbens, G. W. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Avery, R. B., Bostic, R., Calem, P. S., and Canner, G. (1996). Credit risk, credit scoring, and the performance of home mortgages. *Federal Reserve Bulletin*, 82:621–648.

Bhutta, N. and Keys, B. J. (2016). Interest rates and default in the mortgage market. *American Economic Review*, 106(9):2329–2371.

Bhutta, N., Laufer, S., and Rios, I. (2020). Credit scoring and loan performance. *Journal of Financial Economics*, 136(2):388–412.

Campbell, J. Y., Jackson, H. E., Madrian, B. C., and Tufano, P. (2011). Consumer financial protection. *Journal of Economic Perspectives*, 25(1):91–114.

Di Maggio, M., Kermani, A., and Ramcharan, R. (2017). Monetary policy pass-through: Household consumption and voluntary deleveraging. *American Economic Review*, 107(11):4357–4398.

Gerardi, K., Lehnert, A., Sherlund, S. M., and Willen, P. (2008). Making sense of the subprime crisis. *Brookings Papers on Economic Activity*, pages 69–145.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did securitization lead to lax screening? evidence from subprime loans. *Quarterly Journal of Economics*, 125(1):307–362.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press.