

REM WORKING PAPER SERIES

**Out-of-Pocket Health Spending and Income Inequality in the
United States: Threshold Effects and Structural Change**

Selin Ozdamar

REM Working Paper 0437-2026

October 2026

REM – Research in Economics and Mathematics

Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

Any opinions expressed are those of the authors and not those of REM. Short, up to two paragraphs can be cited provided that full credit is given to the authors.





REM – Research in Economics and Mathematics

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912

E-mail: rem@iseg.ulisboa.pt

<https://rem.rc.iseg.ulisboa.pt/>



<https://twitter.com/ResearchRem>

<https://www.linkedin.com/company/researchrem/>

<https://www.facebook.com/researchrem/>

Out-of-Pocket Health Spending and Income Inequality in the United States: Threshold Effects and Structural Change

Abstract

Income inequality across US states has risen to historic highs, and the channels driving it remain incompletely understood. This paper examines a channel that has received limited empirical attention: the financing of health care through out-of-pocket (OOP) payments. Using a balanced panel of 51 US states over 2004–2019, we apply the Hansen (1999) fixed-effect panel threshold model with OOP burden measured relative to household capacity to pay (CTP) and ask whether the relationship between OOP burden and inequality is regime-dependent. We document two findings. First, the marginal inequality effect of OOP burden on Gini-measured income inequality exhibits a statistically significant regime shift at 11.84% of CTP, a level well below the conventional 40% catastrophic-spending threshold, with an effect approximately 17% larger above this level than below. Second, the strength of this channel has not been stable over time: the marginal inequality effect of OOP burden fell by roughly half around 2009, a shift that appears under both burden measures and is better characterized as a discrete break than a gradual decline. We find no additional shift at the 2014 implementation of the Affordable Care Act and no differential decline in Medicaid-expansion states, indicating that the weakening of the channel predates the ACA. These results indicate that the distributional consequences of OOP financing begin to intensify well below the threshold that is defined as catastrophic in the literature and that the channel is time-varying in ways that existing policy-attribution narratives do not capture.

JEL codes: I13, I14, D63, H51

Keywords: out-of-pocket health spending, income inequality, panel threshold model, catastrophic health expenditure, capacity to pay, Affordable Care Act

1. Introduction

Income inequality in the United States has reached its highest level in a century (Frank, 2009; Piketty et al., 2018; Piketty & Saez, 2003), and a growing literature examines the channels through which the structure of the US health system contributes to distributional outcomes. One channel that has received comparatively little empirical attention is the financing of health care through out-of-pocket (OOP) payments. The United States is the largest health spender in the world, devoting roughly 17 percent of GDP to health care over the past decade (CMS National Health Expenditure Accounts). A substantial share of this spending falls directly on households as deductibles, copayments, coinsurance, and costs for uncovered services that bypass the risk-pooling functions of insurance. This financing structure is not evenly distributed across households. Lower-income households spend a larger share of their disposable income on OOP health costs than higher-income households (Ketsche et al., 2011), and the resulting incidence operates as a regressive deduction from household resources that reduces the median income of the poorest decile by an estimated 48 percent, compared with under 3 percent for the wealthiest decile (Christopher et al., 2018).

This paper asks whether the relationship between OOP health spending burden and income inequality is regime dependent. In other words, whether the marginal effect of OOP burden on inequality changes once burden crosses a critical level. The international financial-protection literature has long emphasized catastrophic-spending thresholds defined at fixed shares of household resources, but these cutoffs are normative monitoring conventions rather than levels identified from data. Whether the data indicate a level at which the relationship between OOP burden and inequality qualitatively changes, and where that level falls in the institutional context of US health financing, remains an open empirical question. We address this question on a balanced panel of 51 US states over 2004–2019, applying a fixed-effect panel threshold model

with OOP burden measured against household capacity to pay, and complement the threshold analysis with a structural-break specification around the implementation of the Affordable Care Act (ACA) in 2014.

State level analysis is suitable for this question, for two reasons. First, income inequality in the United States is meaningfully measured and analyzed at the state level. State-level inequality dynamics can diverge in direction from the national trend, so that the state income distribution is not merely a scaled-down version of the national one (Smith, 2021). Also, state-level Gini panels are an established framework for studying the determinants of US income inequality (Ashby & Sobel, 2007; Oyekola, 2021). Second, the policy environment relevant to both inequality and health financing is set at the state level. State redistributive and labor-market policies significantly shape state income distributions (Hatch & Rigby, 2015), and the institutional determinants of OOP exposure (i.e., the administration and generosity of Medicaid, the regulation of insurance markets, and the structure of the individual marketplace) vary across states. Our identification exploits within-state variation in OOP burden over time, holding fixed time-invariant state characteristics and common national shocks, and asks whether the marginal effect of that variation on inequality differs across burden regimes.

We document two complementary findings. First, the effect of OOP health spending on Gini-measured inequality exhibits a modest but statistically significant regime shift at a critical OOP burden of 11.84% of capacity to pay that is under the catastrophic cutoff of the international financial-protection literature. Below this level, a one-unit increase in OOP burden is associated with a 0.797-point increase in the state-level Gini coefficient; above it, the corresponding effect is 0.933, approximately 17 percent larger. Second, the strength of this channel changed markedly within our sample period. The marginal inequality effect of OOP burden fell from approximately

1.60 to 0.87 around 2009, a decline that appears under both the capacity-to-pay and budget-share measures and is better described as a discrete break at the Great Recession than as a gradual trend. We find no additional break at the 2014 implementation of the Affordable Care Act and no differential decline in Medicaid-expansion states, indicating that the weakening of the channel predates the ACA. Taken together, these findings indicate that the distributional consequences of OOP financing begin to intensify well before households reach the conventional catastrophic-spending threshold, and that the strength of this channel varies over time in ways not explained by the timing of the ACA.

The paper makes three contributions. First, we provide the first empirical identification of a threshold in the OOP-burden–inequality relationship at state level, showing that the distributional effect of OOP financing intensifies above a data-determined burden level rather than varying smoothly or only at the extreme tail. Second, we show that this empirical threshold falls substantially below the normative catastrophic-spending cutoffs used to monitor financial protection, implying that conventional headcount measures understate the inequality channel operating in the broad range of moderate OOP burden. Third, we provide evidence that the OOP-inequality channel is time-varying, and that its weakening dates to the Great Recession rather than to the ACA. Microsimulation studies conclude that the ACA reduced US income inequality (Buettgens et al., 2021). Our results add an important qualification to this conclusion: the sensitivity of state-level inequality to OOP burden had already declined before the law's coverage provisions took effect, so a simple before-and-after comparison would overstate the role of the reform.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the data and the econometric framework. Section 4 reports the

main results, and Section 5 examines robustness. Section 6 discusses the mechanisms underlying the threshold finding and the policy implications of the analysis. Section 7 concludes.

2. Financial Protection, Out-of-Pocket Spending and Inequality

This section locates our paper at the intersection of three strands of research: the literature on the conceptual and empirical measurement of household financial protection from out-of-pocket (OOP) health spending; the smaller body of work examining the consequences of OOP spending for income inequality, particularly in the United States; and a recent methodological literature applying panel threshold methods to the relationship between OOP spending and household-level outcomes. Section 2.1 reviews the first strand; Section 2.2 covers the second and methodological strand together; Section 2.3 states the gap our paper addresses.

2.1 OOP Health Spending and Household Financial Protection

The financial-protection literature holds that out-of-pocket health expenditures fall disproportionately on lower-income households, producing financial distress beyond certain thresholds. Wagstaff and van Doorslaer (2003) provide the foundational treatment, defining catastrophic spending as the share of households whose OOP spending exceeds a fixed proportion of total consumption (the budget-share approach) and showing that the choice of denominator materially affects the measurement of incidence. Xu et al. (2003) introduce the capacity-to-pay (CTP) denominator, consumption net of basic-needs spending, and the widely used 40% threshold, on the grounds that CTP better captures the resources households can plausibly redirect toward health spending. Comparing four methods across 14 European countries, Cylus et al. (2018), show that the budget-share approach misrepresents the distribution of hardship across incomes, motivating the CTP denominators adopted here.

The CTP framework anchors two subsequent developments relevant here. Wagstaff et al. (2018), using household survey data from 133 countries, document that catastrophic OOP spending affects hundreds of millions of people worldwide and has been rising over time, providing the empirical basis for the financial-protection indicator embedded in the Sustainable Development Goals (SDG 3.8.2). The SDG indicator monitors the share of households whose OOP spending exceeds 10% or 25% of total consumption, sitting alongside the older 40% capacity-to-pay threshold from Xu et al. (2003). Nguyen et al. (2023) review estimation methods and note that these thresholds are normative rather than data driven and highlight the unresolved problem of identifying empirically the level at which OOP burdens become catastrophic.

Two features of this literature are worth highlighting because they shape our empirical approach. First, the outcomes of interest are predominantly poverty-incidence variables (such as catastrophic-spending headcounts, impoverishment headcounts, or the severity of poverty) rather than distributional measures across the full income range. Second, the burden thresholds are normative rather than empirically identified: the 10%, 25%, and 40% cutoffs are useful monitoring conventions but are not derived from the data.

2.2 OOP Spending, Income Inequality, and Threshold Methods

A smaller and more recent strand of the literature examines the distributional consequences of OOP financing across the full income distribution rather than at the bottom tail. The most direct US-focused contribution is (Christopher et al., 2018), who use Current Population Survey data for 2010–2014 to compute the Gini coefficient before and after subtracting household medical outlays (including OOP and insurance premiums) from income. They find that medical outlays raise the US Gini coefficient by approximately 1.4 percentage points and reduce the median income of the poorest decile by 47.6%, compared with only 2.7%

for the wealthiest decile, demonstrating that OOP financing operates as a regressive deduction from household resources. Ketsche et al. (2011) complement this finding with broader evidence on the regressive incidence of US health-care financing. Lower-income families contribute a greater proportion of their income to national health-care expenditures compared to higher-income families. This trend has remained consistent across various measurement approaches and survey iterations.

If OOP financing is regressive at the household level, the natural next question is whether major reforms to US health insurance modify this distributional pattern. The Affordable Care Act, implemented in 2014, provides a particularly informative test case: it expanded Medicaid in participating states, established income-targeted subsidies through the insurance Marketplaces, and imposed annual cost-sharing limits on covered plans. Goldman et al. (2018) document that OOP spending and premium contributions declined significantly after ACA implementation, with effects concentrated among low-income adults. Within this strand, several papers examine the Affordable Care Act's distributional consequences specifically. Buettgens et al. (2021) the closest substantive parallel to our structural-stability analysis, use microsimulation to compare US income inequality under the ACA with a no-ACA counterfactual, finding a 10.6% reduction in the Theil index, concentrated in Medicaid-expansion states. Liu et al. (2021) show in a difference-in-differences design that marketplace cost-sharing reductions lowered OOP spending and catastrophic-expenditure incidence among eligible low-income enrollees, with little effect on middle-income enrollees which reflects the income-targeted structure of the policy. A related quasi-experimental literature evaluates the ACA's coverage provisions directly. Courtemanche et al. (2019) find that the reform, once fully implemented for three years, reduced the pre-reform disparity in coverage across income groups by about 43%, attributing the whole effect to

Medicaid expansion. Household-level gains of this kind, however, need not alter the aggregate state-level relationship between OOP burden and income inequality, which is the object of our structural-stability analysis. These papers provide evidence on the ACA's distributional role through simulation and quasi-experimental designs, but they do not characterize the regime structure of the underlying OOP–inequality relationship, nor do they test on observed data whether the timing of changes in that relationship aligns with the reform.

The methodologically closest paper to ours is Sirag and Nor (2021), who estimate a dynamic panel threshold model of the poverty effects of OOP spending across 145 countries, using the OOP share of total health expenditure as the threshold variable. Our design differs on each of these dimensions: we estimate the static Hansen (1999) fixed-effect threshold model on a panel of 51 US states, with the Gini coefficient as the outcome and the capacity-to-pay measure as the threshold variable.

2.3 Contribution

Three gaps in the existing literature motivate our paper. First, the financial-protection literature relies on normative thresholds (10%, 25%, 40%) rather than empirically identified ones; the level at which OOP burden begins to generate hardship-grade distributional consequences in a specific institutional context has not been estimated. Second, the literature on OOP and inequality in the US either uses national-level decompositions (Christopher et al., 2018) or microsimulation (Buettgens et al., 2021), without empirically identifying a threshold in the burden–inequality relationship or characterizing its regime structure on observed state-year data. Third, the small body of panel-threshold work on OOP spending focuses on cross-country panels and poverty outcomes; no existing study applies these methods to US state-level data with an inequality outcome.

Our paper addresses these gaps along three dimensions. We apply the Hansen (1999) fixed-effect panel threshold model to a balanced panel of 51 US states over 2004–2019, identifying the threshold in the OOP-burden–inequality relationship empirically rather than imposing it normatively, and find a level substantially lower than the cutoffs used in financial-protection monitoring frameworks. We adopt the CTP framework of Xu et al. (2003) and Cylus et al. (2018) as the basis for measuring OOP burden, in contrast to the OOP-to-total-health-expenditure ratio used in the prior threshold work (Sirag & Nor, 2021). And we examine the structural stability of the OOP-inequality relationship over time, testing whether its evolution can be attributed to the ACA using both the timing of the reform and cross-state variation in Medicaid expansion.

3. Data and Methodology

This section describes the data, variable construction, and econometric framework used to test whether out-of-pocket (OOP) health spending exhibits a regime-dependent effect on income inequality across US states. Section 3.1 presents the data sources and the construction of the OOP burden measure. Section 3.2 develops the Hansen (1999) fixed-effect panel threshold model that constitutes the main empirical strategy. Section 3.3 outlines robustness specifications used to assess the credibility of the main results.

3.1 Data and Variable Construction

The empirical analysis is conducted on a balanced panel of 51 US states (the 50 states plus the District of Columbia) over the period 2004–2019, yielding 816 state-year observations. The sample period begins in 2004 because consistent state-level health expenditure data from the Institute for Health Metrics and Evaluation (IHME) become available from this year and ends in 2019 to avoid the structural disruption introduced by the COVID-19 pandemic on both health

spending patterns and labor-market measures of inequality. Table 1 reports descriptive statistics for all variables. The full panel variation decomposition, distinguishing between-state and within-state variation, is reported in Table A1 of appendix A.

The dependent variable is the state-level Gini coefficient of pre-tax income inequality, the most widely used scalar measure of income inequality (Atkinson, 1970). We retrieved our Gini variable from the updated Frank (2009) state-level inequality series, which provides a consistent annual measure across US states based on Internal Revenue Service (IRS) administrative tax microdata on adjusted gross income. AGI is a broad income concept that includes wages and salaries, capital income (dividends, interest, rents, and royalties), and entrepreneurial income, but excludes interest on state and local bonds and transfer income from federal and state governments. Because IRS data are reported in grouped income brackets, Frank constructs the Gini coefficient using the compromise interpolation method of Cowell & Mehta (1982), which combines lower-bound and upper-bound bracket-based estimates of the Gini. The Frank measure has the advantage of being computed directly from administrative records rather than survey data, ensuring greater comparability across states and over time, and the state-level series correlates closely with the aggregate Piketty & Saez (2003) US time series of top income shares. The principal limitation, common to all IRS-based inequality measures, is that individuals with income below the tax filing threshold are omitted; this introduces some attenuation of bottom-tail inequality. Over our sample period, the mean state-level Gini is 0.607, with state-year values ranging from 0.536 to 0.720.

The central explanatory variable is OOP health spending burden, measured under the capacity-to-pay (CTP) approach proposed by (Xu et al., 2003) and refined by Cylus et al. (2018).

Under this approach, OOP burden is defined as the ratio of OOP health expenditure to capacity to pay, where capacity to pay is total household consumption net of basic-needs consumption (food, clothing, and housing including utilities). Formally, for state i in year t :

$$OOP_burden_{it} = OOP_{pc_{it}} / (Consumption_{pc_{it}} - BasicNeeds_{pc_{it}}) \quad (1)$$

where all variables are expressed in per capita terms. The CTP approach is preferred over the simpler budget-share approach (OOP divided by total consumption) on theoretical grounds: by deducting basic-needs spending from the denominator, the CTP measure isolates the discretionary income that households can plausibly redirect toward health expenditures, providing a more meaningful indicator of the financial hardship imposed by OOP spending (Cylus et al., 2018; Nguyen et al., 2023). Households with limited discretionary income, such as those in lower-income states or with high housing costs, will exhibit higher CTP burden for the same nominal OOP outlay, whereas the budget-share measure obscures this heterogeneity. We re-estimate the main specification under the budget-share approach as a robustness check, reported in Section 5.

OOP and total health expenditure data are drawn from the IHME State-Level Health Expenditure database, which decomposes annual state-level health spending by payer (OOP, private insurance, Medicare, Medicaid, and other public sources). State-level personal consumption expenditure and basic-needs consumption (food and beverages purchased for off-premises consumption, clothing and footwear, and housing and utilities) come from the Bureau of Economic Analysis (BEA) Personal Consumption Expenditure by State series. Real GDP per capita (in chained 2017 dollars) is also from BEA. Annual state population estimates are taken from the Federal Reserve Economic Data (FRED) repository. All nominal variables are converted to per capita units using contemporaneous population estimates.

Three additional explanatory variables enter the model as regime-independent controls. The natural logarithm of real GDP per capita captures the macroeconomic environment of each state. The public health spending share, defined as the sum of Medicare and Medicaid expenditures divided by total health expenditure, captures the extent of public coverage. The private insurance share, defined as private insurance spending divided by total health expenditure, captures the role of employer-sponsored and individually purchased coverage.

3.2 Empirical Strategy

3.2.1 Baseline linear specification

We begin with a baseline two-way fixed-effects model that constrains the relationship between OOP burden and income inequality to be linear:

$$Gini_{it} = \mu_i + \delta_t + \beta \cdot OOP_burden_{it} + X_{it}'\gamma + \varepsilon_{it} \quad (2)$$

where μ_i denotes state fixed effects absorbing time-invariant unobserved heterogeneity (e.g., demographic composition, political culture, geography), δ_t denotes year fixed effects absorbing aggregate shocks common to all states (e.g., the 2008 financial crisis, federal policy changes), X_{it} is the vector of regime-independent controls (log GDP per capita, public spending share, private spending share), and ε_{it} is the error term. The coefficient β summarizes the average marginal effect of a one-unit increase in OOP burden on the Gini coefficient. A Hausman (1978) test rejects the null that random effects is consistent ($\chi^2 = 12.50$, $p = 0.014$), confirming the appropriateness of the fixed-effects estimator.

For all linear specifications we report (Driscoll & Kraay, 1998) standard errors, which are robust to cross-sectional dependence, serial correlation, and heteroscedasticity in panels with small T. The maximum lag for the Newey-West kernel is set to three following Hoechle (2007). The supporting residual diagnostics are reported in Section 4.1.

We supplement the linear specification with a quadratic term in OOP burden to test for smooth nonlinearity. A failure to reject the null that the squared term is zero would not preclude regime-dependent nonlinearity, in which the marginal effect of OOP burden changes across regimes defined by the level of OOP burden itself rather than along a smooth curve. This motivates the threshold approach developed in Section 3.2.2.

3.2.2 Hansen (1999) panel threshold model

The central econometric framework of the paper is the fixed-effect panel threshold model developed by Hansen (1999) and implemented in Stata by Wang (2015). Unlike polynomial specifications, the threshold model accommodates non-smooth, regime-switching relationships in which the marginal effect of a regressor changes abruptly when an observable variable crosses an endogenously determined threshold.

The single-threshold specification takes the form:

$$Gini_{it} = \mu_i + \beta_1 \cdot OOP_burden_{it} \cdot I(OOP_burden_{it} < \gamma) + \beta_2 \cdot OOP_burden_{it} \cdot I(OOP_burden_{it} \geq \gamma) + X_{it}'\delta + \varepsilon_{it} \quad (3)$$

where $I(\cdot)$ denotes the indicator function, equal to one when the condition in parentheses is satisfied and zero otherwise, γ is the unknown threshold parameter, and OOP burden serves jointly as the regime-dependent variable (whose coefficient changes between regimes) and the threshold variable (which determines regime membership). The coefficients β_1 and β_2 capture the marginal effect of OOP burden on inequality below and above the threshold respectively. Year dummies are included among the regime-independent controls but absorbed in X_{it} to keep the equation readable. The threshold variable enters in levels rather than in transformations to preserve the economic interpretation of the threshold value.

Estimation proceeds by concentrated least squares. For each candidate threshold value γ in a fine grid over the empirical support of OOP burden (we use 400 grid points, trimming 5% from each tail to ensure each regime contains a sufficient number of observations), the within-transformed model is estimated by OLS, and the residual sum of squares $S(\gamma)$ is computed. The threshold estimate is the value that minimizes this concentrated criterion:

$$\hat{\gamma} = \arg \min_{\gamma} S(\gamma) \quad (4)$$

Confidence intervals for γ are constructed by inverting the likelihood ratio statistic $LR(\gamma) = [S(\gamma) - S(\hat{\gamma})] / \hat{\sigma}^2$, where $\hat{\sigma}^2$ is the estimated residual variance. The 95% confidence region is the set of values for which $LR(\gamma)$ does not exceed the 7.35 critical value from Hansen (1999). Following (Wang, 2015), we note that these intervals can be artificially narrow in finite samples and interpret their precision with caution.

The existence of a threshold effect is tested formally via the bootstrap procedure proposed by Hansen (1999). Under the null hypothesis that no threshold exists ($\beta_1 = \beta_2$), the threshold parameter γ is unidentified, so standard inference does not apply. The test statistics

$$F = (S_0 - S(\hat{\gamma})) / \hat{\sigma}^2 \quad (5)$$

compares the residual sum of squares from a linear model (S_0) with that from the threshold model ($S(\hat{\gamma})$). Hansen shows that bootstrap critical values can be obtained by repeatedly drawing samples from the within-transformed residuals, re-estimating the model, and computing the F statistic on each replicate. We use 2,000 bootstrap replications throughout the main results to obtain stable p-values.

We follow Hansen's (1999) sequential testing procedure to determine the appropriate number of regimes. Conditional on rejecting the null of no threshold, we test whether a second threshold exists; conditional on rejecting the no-second-threshold null, we test for a third. The procedure stops at the highest-order specification for which the additional threshold is statistically significant. For our main CTP specification, this procedure retains the single-threshold model.

The Hansen (1999) estimator requires the threshold variable to be stationary. We test this requirement using both the Fisher-type augmented Dickey-Fuller test (Choi, 2001) and the cross-sectionally robust Pesaran (2007) CADF test; results are reported in Section 4.1.

3.2.3 Structural Stability and the ACA

The threshold model identifies regime-switching in the OOP burden–inequality relationship without specifying what generates the regime structure. To investigate whether the implementation of the Affordable Care Act (ACA) in 2014 (the Medicaid expansion and individual marketplace reforms it introduced) modified this relationship, we estimate a complementary specification that augments the linear fixed-effects model with an interaction between OOP burden and a post-ACA indicator:

$$Gini_{it} = \mu_i + \delta_t + \beta_1 \cdot OOP_burden_{it} + \beta_2 \cdot (OOP_burden_{it} \times Post_t) + X_{it}'\gamma + \varepsilon_{it} \quad (6)$$

where $Post_t$ is an indicator equal to one for years 2014 and later. Under this specification, β_1 captures the marginal effect of OOP burden on inequality during the pre-ACA period, while $\beta_1 + \beta_2$ gives the implied post-ACA effect. Because a single-break specification imposes the break date rather than testing it, we treat the ACA date as one candidate among many. We estimate the

interaction model at every candidate break year from 2007 to 2017, compare the preferred break against a smooth linear decline in the marginal effect, and estimate a nested specification in which the leading break and the 2014 date enter jointly. Finally, we exploit cross-state variation in Medicaid expansion: if the ACA moderates the channel, the decline should be concentrated in expansion state-years. Expansion timing is taken from the Kaiser Family Foundation expansion tracker¹, coded as active from year t if coverage began on or before July 1 of year t and from $t+1$ otherwise. We estimate equation (6) under both the CTP and budget-share measures of OOP burden, with Driscoll-Kraay standard errors throughout.

4. Results

This section presents the main empirical findings. Section 4.1 reports the diagnostic tests that motivate the inferential choices laid out in the methodology. Section 4.2 establishes the baseline linear effect of OOP burden on income inequality and demonstrates the absence of smooth nonlinearity, motivating the threshold approach. Section 4.3 reports the central result of the paper: the Hansen (1999) panel threshold model identifies a critical OOP burden level above which the marginal inequality effect of OOP spending strengthens. Section 4.4 examines whether the implementation of the Affordable Care Act in 2014 modified this relationship. Section 4.5 synthesizes the findings.

¹ Kaiser Family Foundation produced a dataset called “Status of State Action on the Medicaid Expansion Decision” which can be found at: <https://www.kff.org/affordable-care-act/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>

4.1 Diagnostic Tests

Three standard diagnostic tests applied to the residuals of the linear fixed-effects model in equation (2) confirm the presence of the residual structure anticipated in Section 3. The results of diagnostic tests are represented in the table 2. The Pesaran (2004) CD test rejects the null of cross-sectional independence ($CD = 27.87, p < 0.001$), with an average absolute pairwise correlation of 0.567 across states which is a magnitude consistent with the prevalence of national shocks affecting all US states simultaneously. The Wooldridge (2010) test rejects the null of no first-order autocorrelation ($F = 41.32, p < 0.001$), reflecting the well-documented persistence of state-level inequality. The modified Wald test for groupwise heteroscedasticity strongly rejects the null of equal error variances across states ($\chi^2 = 4,689, p < 0.001$), consistent with the substantial heterogeneity in state population sizes and economic structures. These rejections jointly justify the use of Driscoll-Kraay (1998) standard errors for all linear specifications in the analysis.

Panel unit root tests support the stationarity assumption underlying the Hansen (1999) threshold estimator. The threshold variable (OOP burden under the CTP measure) strongly rejects the unit root null under the cross-sectionally robust Pesaran (2007) CADF test ($t\text{-bar} = -3.14, p < 0.01$), satisfying the central stationarity requirement of the threshold framework. The Gini coefficient and the public spending share also reject the unit root null under Fisher-type ADF tests. Log GDP per capita and the private spending share fail to reject in some specifications, but this likely reflects the limited power of panel unit root tests with $T = 16$ (Pesaran, 2007). All series are bounded by construction (Gini and financing shares lie in the unit interval) or display deterministic rather than stochastic trends, and the within-state transformation applied by the fixed-effects estimator absorbs persistent state-specific level differences.

4.2 Baseline Linear Effects

Table 3 reports the baseline linear and quadratic fixed-effects models. Column (1) presents the linear specification with cluster-robust standard errors. The coefficient on OOP burden is positive and marginally significant ($\beta = 1.169$, $p = 0.034$), indicating that a one-unit increase in OOP burden as a share of capacity to pay is associated with a 1.17-point increase in the Gini coefficient. Both the public spending share ($\beta = 0.613$, $p = 0.007$) and the private spending share ($\beta = 0.524$, $p = 0.020$) enter positively and significantly, while log GDP per capita is statistically indistinguishable from zero within states.

Column (3) re-estimates the linear specification using Driscoll-Kraay standard errors. The point estimates are identical by construction, but the corrected inference is markedly stronger: the coefficient on OOP burden is significant at the 1% level ($p = 0.005$), and the spending-share controls become significant at the same level. This pattern, in which Driscoll-Kraay inference strengthens rather than weakens the linear effects, reflects the fact that cluster-robust standard errors can be inefficient relative to Driscoll-Kraay when cross-sectional correlation in the residuals is positive. Throughout the remainder of the paper, we report Driscoll-Kraay inference for linear specifications and the default bootstrap inference of Wang (2015) for the threshold model.

Column (2) tests for smooth nonlinearity by adding a quadratic term in OOP burden. The squared term is positive but statistically insignificant ($\beta = 5.569$, F-test $p = 0.178$), and the linear term collapses to economic and statistical insignificance once the quadratic is included. Two interpretations are consistent with this finding. First, the relationship between OOP burden and

inequality may genuinely be linear within states. Second, the relationship may exhibit a structural break that a continuous polynomial cannot accommodate, in which case the within-state effect of OOP burden differs across regimes defined by the level of OOP burden itself rather than along a smooth curve. Section 4.3 tests this hypothesis directly.

4.3 Hansen Panel Threshold Model

Table 4 reports the main result of the paper: the Hansen (1999) single-threshold panel model under the CTP specification. The Hansen sequential test for the existence of a threshold strongly rejects the null of no threshold against the single-threshold alternative ($F = 34.14$, bootstrap $p = 0.034$, 2,000 replications), exceeding the 5% bootstrap critical value of 30.34. Under the same procedure, the test of the single-threshold null against a double-threshold alternative fails to reject ($F = 17.70$, bootstrap $p = 0.195$), confirming that the data favors a single critical OOP burden level rather than a richer regime structure. The single-threshold specification is therefore retained as the preferred model. This bootstrap procedure is also the formal test of regime-coefficient equality: under the null hypothesis $\beta_1 = \beta_2$ the threshold parameter is not identified, so the F_1 statistic is evaluated against bootstrap rather than standard critical values. A complementary Wald test on the regressor split at the estimated threshold likewise rejects equality ($F(1, 15) = 14.77$, $p = 0.002$).²

The estimated threshold lies at $\hat{\gamma} = 0.1184$ of capacity to pay, that is when out-of-pocket health spending reaches 11.84% of disposable income net of basic needs. The 95% confidence interval for the threshold, constructed via inversion of the likelihood ratio statistic, is [0.1182,

² This test conditions on the estimated threshold of 0.1184 rather than accounting for its sampling variation, so we treat it as corroborating the bootstrap test rather than replacing it. The auxiliary split regression yields slopes of 0.806 and 0.938, close to the threshold-model estimates.

0.1185]. The narrowness of this interval should be interpreted with caution: Wang (2015) notes that LR-inversion confidence intervals can be artificially narrow in finite samples, particularly when the likelihood surface has a single sharp minimum, as ours does. The likelihood profile underlying this interval is shown in Figure 1; the V-shape of the LR statistic centered at $\hat{\gamma}$ confirms that the threshold is well-identified and not the result of a flat or multimodal likelihood surface.

Regime-dependent coefficients reveal the structure of the threshold relationship. Below the threshold (state-years with CTP burden under 11.84%, comprising the bulk of the sample), a one-unit increase in OOP burden is associated with a 0.797-point increase in the Gini coefficient ($p = 0.001$). Above the threshold, the corresponding coefficient is 0.933 ($p < 0.001$). The regime difference (i.e., $\beta_2 - \beta_1 = 0.136$) implies that the marginal inequality effect of OOP burden is approximately 17% larger above the threshold than below it. The regime-independent controls retain stable signs and significance: the public spending share ($\beta = 0.520$, $p < 0.001$) and private spending share ($\beta = 0.467$, $p < 0.001$) both raise inequality, while log GDP per capita is again statistically insignificant within states. The within-state R^2 rises from 0.173 in the linear specification to 0.207 in the threshold model, reflecting the additional explanatory power provided by allowing the OOP burden coefficient to vary across regimes. The high-burden regime is substantial in size: 153 of 816 state-year observations (18.8%) lie above the estimated threshold. The share of state-years above the threshold is 24.7% in 2004–2008, 22.0% in 2009–2013, and 11.1% in 2014–2019, so both regimes are populated throughout the sample, and the threshold does not repackage a time effect.

Several features of this result deserve emphasis. First, the regime difference is modest: a 17% increase in the marginal effect, a kink rather than a sharp switch. Second, both regime

coefficients are positive and significant, so the threshold modulates the strength of the inequality reducing effect rather than its direction. Third, the bootstrap p-value of 0.034 clears the 5% level without being extreme; replication in other samples would be valuable.

Figure 2 provides a visual depiction of the regime-dependent relationship. The figure plots the partial regression of the Gini coefficient on OOP burden after residualizing both variables on state fixed effects, year fixed effects, log GDP per capita, public spending share, and private spending share. Linear fit lines are estimated separately for state-years with OOP burden below and above the estimated threshold. After partialling out controls and fixed effects, both fits slope upward, consistent with the positive regime coefficients reported in Table 3. The figure helps to clarify the structure of the threshold result: the regime difference identified by the Hansen estimator reflects a steeper within-state response of inequality to OOP burden among state-years that have crossed the threshold, not a discontinuous level shift in inequality at the threshold itself.

Substantively, the estimated threshold of 11.84% of capacity to pay is striking when set against the conventional 40% catastrophic spending threshold proposed by Xu et al. (2003) and used as the reference cutoff in much of the international financial protection literature. Our results suggest that the inequality consequences of OOP spending begin to intensify well below the catastrophic threshold, i.e., at a level of OOP burden that is unremarkable in the cross-country distribution. This finding implies that exclusive focus on catastrophic-spending headcounts may understate the distributional consequences of OOP financing, particularly in mature health systems with already-substantial private cost-sharing.

4.4 Structural Stability of the OOP-Inequality Relationship

The threshold model identifies regime-switching in the OOP burden–inequality relationship without specifying what drives the regime structure. A natural candidate is the Affordable Care Act (ACA), implemented in 2014, which fundamentally altered the structure of US health financing through Medicaid expansion in participating states, the establishment of subsidized individual marketplaces, and the introduction of essential-benefit and cost-sharing standards. To examine whether the ACA modified the inequality channel of OOP spending, we estimate equation (6), which augments the linear fixed-effects model with an interaction between OOP burden and a post-2014 indicator.

Table 5 presents the results under both the CTP and budget-share measures of OOP burden, with Driscoll-Kraay standard errors throughout. Columns 1 and 2 reproduce the conventional specification with a break at ACA implementation under the CTP and budget-share measures: the interaction is negative and statistically significant in both (-0.426 , $p = 0.032$; -0.548 , $p = 0.041$), which taken alone would suggest a post-2014 weakening of the channel. The remaining columns show that this reading misdates the shift. Scanning all candidate break years from 2007 to 2017 (Appendix A, Figure A1), the interaction is largest and most precisely estimated at 2009 under both measures (columns 3 and 4: -0.799 , $t = -7.78$ for CTP; -1.075 , $t = -7.71$ for BS), with the test statistic declining monotonically thereafter and no secondary peak in 2014.³ The discrete 2009 break also fits better than a linear decline in the marginal effect (within R^2 0.321 against 0.290) and improves markedly on the 2014 specification (0.321 against 0.231). When the 2009

³ Because the break date is selected by scanning eleven candidate years, conventional critical values for a single prespecified break date do not apply (Andrews, 1993). The relevant benchmark is the sup-F critical value for a break at an unknown date, approximately 9 at the 5% level for one restricted parameter with 15% trimming. The peak statistic at 2009 corresponds to F of approximately 60, so the break survives any correction of this form.

and 2014 breaks enter jointly (column 5), the Recession-era break is essentially unchanged ($-0.731, p < 0.001$) while the additional 2014 increment is small and marginally significant ($-0.115, p = 0.092$). Column 6 tests the cross-sectional implication of an ACA effect: the decline shows no concentration in Medicaid-expansion state-years ($-0.084, p = 0.251$), while the 2009 break is undiminished.

Figure 3 plots the marginal effect of OOP burden by year, from a specification interacting the burden measure with the full set of year indicators. The estimates form two plateaus, near 1.5 through 2008 and near 0.7 from 2010 onward, with the decline concentrated at 2009 and 2010 and no further shift at 2014.

Two remarks qualify these results. First, the ACA's nationally uniform provisions, including the annual cost-sharing limit and essential health benefits requirements, applied to all states simultaneously and therefore cannot be separated from the Recession-era break using state-level variation; the evidence rules out the detectable state-level implications of an ACA effect rather than every conceivable channel. Second, the secular flattening of the channel works against, not for, the threshold finding: high-burden state-years are concentrated late in the sample when the marginal effect is smaller, yet the threshold model nonetheless identifies the steeper slope in the high-burden regime.

4.5 Synthesis

Three findings emerge from the analysis. First, OOP health spending burden has a positive within-state effect on income inequality across US states, robust to alternative inference procedures and to the inclusion of standard macroeconomic and health-financing controls (Section 4.2). Second, this effect is regime-dependent: the marginal inequality effect of OOP

burden is approximately 17% larger when CTP burden exceeds 11.84% of capacity to pay than below this critical level (Section 4.3). Third, the strength of the channel is time-varying: the marginal inequality effect of OOP burden fell by roughly half around 2009 under both burden measures, with no detectable additional shift at ACA implementation and no differential decline in Medicaid-expansion states (Section 4.4).

5. Robustness

We assess the robustness of the main findings along five dimensions: excluding an outlier state, re-estimating the threshold model under an alternative measure of OOP burden, exploring whether a higher-order threshold structure exists under that alternative measure, controlling for labor-market conditions, and re-estimating the model with alternative inequality measures. Results are reported in Tables 6 and 7.

5.1 Excluding the District of Columbia

The District of Columbia is an outlier on real GDP per capita (Figure 4; more than three standard deviations above the cross-state mean) and differs structurally from the fifty states in demographic composition and the concentration of federal employment. To verify that DC is not driving the threshold result, columns (1) and (2) of Table 6 re-estimate the single-threshold model excluding DC, under the CTP and budget-share measures of OOP burden respectively.

The threshold estimates are essentially unchanged. Under the CTP measure, the threshold moves from $\hat{\gamma} = 0.1184$ in the full sample to $\hat{\gamma} = 0.1187$ without DC, a difference of three basis points. Under the budget-share measure, the threshold moves from 0.0528 to 0.0533. The bootstrap evidence in favor of a threshold strengthens in both specifications: the bootstrap p-value falls from 0.034 to 0.027 under CTP and from 0.078 to 0.034 under the budget-share

measure. The regime-dependent coefficients retain their qualitative pattern: under CTP, the marginal effect remains larger above the threshold than below ($\beta_2 = 0.793$ versus $\beta_1 = 0.660$), with the regime difference of approximately 20% close to the 17% reported in the full sample. We conclude that the threshold result is not driven by the inclusion of DC.

5.2 Alternative Measure of OOP Burden

Our main specification uses the capacity-to-pay measure of OOP burden, which deducts basic-needs consumption from the denominator following Xu et al. (2003) and Cylus et al. (2018). The simpler budget-share measure proposed by Wagstaff and van Doorslaer (2003), which uses total personal consumption as the denominator, remains widely used in the financial-protection literature and is the basis for the SDG 3.8.2 catastrophic-spending indicator. Column (2) of Table 6 re-estimates the single-threshold model under the budget-share measure on the sample excluding DC.

The estimated threshold under the budget-share measure is $\hat{\gamma} = 0.0533$, considerably lower than the CTP threshold because the denominator is mechanically larger (total consumption rather than discretionary capacity to pay). Despite the difference in scale, the qualitative finding is preserved: the bootstrap test rejects the linear null at the 5% level ($p = 0.034$), and the marginal effect of OOP burden on inequality is positive and statistically significant in both regimes ($\beta_1 = 1.301$ below threshold, $\beta_2 = 0.885$ above). The regime ordering under the budget-share measure differs from the CTP specification (i.e., the marginal effect is larger below the threshold than above), which would suggest a saturation pattern in which incremental burden has more inequality consequences when overall burden is low. The next subsection examines whether this

apparent ordering reversal reflects a richer regime structure not captured by the single-threshold specification.

5.3 Higher-Order Threshold Specification

Column (3) of Table 6 reports the double-threshold specification under the budget-share measure on the full sample. Hansen's sequential testing procedure rejects the single-threshold null in favor of the double-threshold alternative ($F_2 = 21.23$, bootstrap $p = 0.091$, marginal at the 10% level) but does not reject the double-threshold null in favor of a triple-threshold model ($F_3 = 4.57$, $p = 0.917$). The data thus support a two-regime structure under the budget-share measure but not a richer one.

The two estimated thresholds, $\hat{\gamma}_1 = 0.0528$ and $\hat{\gamma}_2 = 0.0835$ (Appendix A, Figure A2), partition the OOP burden distribution into three regimes. The regime-dependent coefficients trace a non-monotonic pattern: $\beta_0 = 1.046$ in the lowest regime, $\beta_1 = 0.691$ in the intermediate regime, and $\beta_2 = 0.841$ in the highest regime and there is a U-shape across regimes. The reversal of regime ordering observed in the budget-share single-threshold specification is therefore imposing only one threshold on a relationship that exhibits non-monotonic regime dependence. The upper threshold of $\hat{\gamma}_2 = 0.0835$ (OOP spending equal to 8.35% of personal consumption) is notable for its proximity to the WHO/SDG 10% catastrophic-spending threshold proposed by Wagstaff et al. (2018), suggesting that the inequality consequences of OOP burden begin to intensify shortly before the conventional catastrophic-spending threshold is reached.

We retain the CTP single-threshold specification as the preferred main result on theoretical grounds: the CTP measure is the more appropriate denominator for studying financial hardship, and the single-threshold structure is the most parsimonious model supported by the data under

that measure. The budget-share double-threshold specification is presented here for completeness and because the upper threshold provides an interesting empirical anchor in relation to the international catastrophic-spending literature.

5.4 Controlling for Labor-Market Conditions

Since the structural break documented in Section 4.4 coincides with the Great Recession, we verify that the results are not driven by labor-market slack. Column (4) of Table 6 re-estimates the threshold model with the state unemployment rate (BLS Local Area Unemployment Statistics, annual averages) added to the control set. The result is essentially unchanged: the estimated threshold remains at 11.84% of CTP (bootstrap $p = 0.025$), and both regime coefficients remain positive and significant (0.614 below, 0.753 above). The break specification behaves the same way: the post-2009 decline is undiminished with unemployment included (-0.792 , $t = -6.77$, against -0.799 without), while unemployment itself enters with a small, at most marginally significant coefficient in the Driscoll-Kraay specifications. The Recession-era break is therefore not an artifact of omitted cyclical conditions.

5.5 Alternative Inequality Measures

The Frank (2009) panel also reports top-sensitive inequality measures, and Table 7 re-estimates the main specifications with the Theil index and the Atkinson index ($\epsilon = 0.5$) as outcomes. The results differ sharply from the Gini: no threshold is detected under either measure, and the within-state association between OOP burden and top-sensitive inequality is negative, attenuating toward zero after 2009. This divergence is informative rather than anomalous. Gini responds most strongly to the middle of the income distribution, whereas the Theil and low-aversion Atkinson indices are dominated by the top tail (De Maio, 2007), which in tax-based data moves primarily

with capital income. The threshold and break results therefore characterize Gini-measured inequality, and the inequality reducing OOP channel appears to operate through the broad middle of the distribution, where the consumption-crowding mechanism binds, rather than through top-tail dynamics. Section 6.3 discusses the measurement implications.

5.6 Synthesis

Across the first four robustness exercises, the central finding of regime-dependent inequality effects of OOP burden is preserved. The threshold under the CTP measure is essentially identical with and without DC, and the bootstrap evidence strengthens in both specifications when DC is removed. The threshold finding survives the alternative budget-share measure of OOP burden, with a sequentially preferred double-threshold specification under that measure that uncovers a richer non-monotonic regime structure. The main results of the paper therefore do not depend on the inclusion of DC, on the choice of OOP burden denominator, on the constraint to a single-threshold specification, or on the omission of cyclical labor-market conditions. The fifth exercise qualifies rather than confirms the main results: the threshold structure and the positive OOP association are specific to the Gini and do not extend to top-sensitive inequality measures.

6. Discussion

6.1 Why Does the Threshold Emerge Around 12% of Capacity to Pay?

Two complementary mechanisms explain why the inequality effect of OOP burden intensifies specifically around 12% of capacity to pay rather than at an arbitrary level: a general consumption-crowding channel and a US-specific insurance cost-sharing channel.

When OOP burden is a small share of discretionary income, households at all income levels absorb it through marginal adjustments to non-essential consumption, and the distributional consequences are modest. As OOP costs rise, lower-income households are pushed to cut back on essentials (i.e., food, education, or savings) at lower spending levels than higher-income ones, widening the gap in how hard each group is hit. The crowding-out effect of OOP on non-health consumption has been shown to be sharply income-graded: in the US, Christopher et al. (2018) estimate that household medical outlays reduce the median income of the poorest decile by 47.6% relative to only 2.7% for the wealthiest decile. Above approximately 12% of CTP, a growing share of US households, disproportionately those in lower-income states, cross from the marginal-adjustment regime into the welfare-relevant-displacement regime, generating the regime shift identified by the threshold model.

The insurance cost-sharing channel operates through the structure of US health financing during 2004–2019 and helps explain why the threshold falls at this specific level. According to the Kaiser Family Foundation Employer Health Benefits Survey, the share of covered workers enrolled in high-deductible plans rose from 4% in 2006 to 30% in 2019, while the average annual deductible for single coverage among workers with a deductible rose from \$533 in 2009 to \$1,655 in 2019 (KFF, 2019). The implications for inequality become clear when these cost-sharing levels are expressed relative to capacity to pay. A \$2,000–\$3,000 annual deductible represents approximately 8–12% of CTP for a household with per-capita discretionary income near the sample mean of \$25,000. The empirically identified threshold of 11.84% sits within this range, suggesting that standard insurance cost-sharing provisions become financially binding for lower-income enrollees at the point where the threshold model identifies a regime shift. Galbraith et al. (2011) document that lower-income families in high-deductible plans were almost twice as

likely as lower-income families in traditional plans to spend more than 3% of income on health care (53% versus 29%), consistent with the binding-constraint interpretation.

The timing of the structural break invites a different set of candidate explanations than the policy narrative would suggest. We emphasize that these are candidate interpretations of an associational finding: two are consistent with the evidence, though our design does not adjudicate between them. First, the Great Recession changed the composition of top incomes: the collapse and slow recovery of capital income compressed the upper tail of state income distributions, altering the sources of Gini variation and plausibly loosening its link to health-financing burden. Second, the recession-era expansion of the safety net, including enrollment growth in existing Medicaid categories and the temporary increase in federal Medicaid matching under the American Recovery and Reinvestment Act, cushioned precisely the lower-income households through which the consumption-crowding channel operates. Both mechanisms predate the ACA's coverage provisions, consistent with the location of the break in 2009. Identifying the mechanism behind the Recession-era break is a task for household-level data.

6.2 Policy Implications

Two policy implications follow. First, the estimated threshold of 12% of capacity to pay lies well below the conventional 40% catastrophic-spending threshold proposed by Xu et al. (2003) and below the 10% household-consumption threshold used in the SDG 3.8.2 indicator (Wagstaff et al., 2018). The inequality consequences of OOP financing therefore begin to intensify long before households reach the level conventionally classified as facing catastrophic spending. Standard measures of catastrophic out-of-pocket health spending only track the worst-hit people, but they miss the financial strain felt by a much larger group who are moderately burdened.

Hence, health financing systems need better tools to measure inequality across all levels of spending that capture distributional consequences below catastrophic-level cutoffs.

Second, the structural-stability results carry a cautionary implication for policy attribution.

Microsimulation evidence indicates that the ACA reduced US income inequality (Buettgens et al., 2021), and household-level studies document that specific ACA provisions lowered OOP burden for targeted groups (Goldman et al., 2018; Gotanda et al., 2020; Liu et al., 2021). Our state-level evidence does not contradict those findings, but it shows that the weakening of the OOP-inequality channel, which a before-and-after comparison would naturally credit to the reform, was already in place by 2010. Assessments of the ACA's distributional consequences that rely on pre/post comparisons risk absorbing this earlier structural change into the estimated policy effect.

6.3 Limitations

Four limitations warrant acknowledgment. First, the threshold model identifies regime dependence in the conditional relationship between OOP burden and inequality but does not establish a causal interpretation of the threshold value or the regime difference. Second, the structural-stability analysis locates a break in 2009 but does not identify the mechanism behind it; and because the ACA's nationally uniform provisions applied to all states simultaneously, their effects cannot be separated from the Recession-era break using state-level variation. Third, the analysis uses state-year aggregates, which limit inference about household-level mechanisms; an analogous analysis using household-level survey data such as the Medical Expenditure Panel Survey would provide direct estimates of the consumption-crowding and cost-sharing channels at the household level and complement the state-level evidence presented here. Fourth, the Frank

(2009) inequality series is constructed from IRS adjusted gross income records. It therefore excludes non-filers, who are disproportionately low-income households and include many of the households central to the consumption-crowding mechanism; it treats realized capital gains inconsistently over time; and it omits transfer income. Because the 2009 break coincides with the collapse of realized capital gains in the Great Recession, part of the estimated decline in the marginal effect may reflect a change in the composition of measured income rather than a change in the transmission from OOP burden to inequality. The alternative-measure estimates in Section 5.5 sharpen this caveat. Top-sensitive indices from the same panel show no threshold and a negative OOP association, so our findings are specific to Gini-measured inequality and to the segment of the distribution the Gini emphasizes. We accordingly interpret the threshold and the break as features of the middle tail of the income distribution rather than of the distribution as a whole.

7. Conclusion

This paper has examined whether the relationship between out-of-pocket health spending burden and income inequality across US states is regime-dependent and whether major insurance reforms modify this relationship. Using a balanced panel of 51 US states over 2004–2019 and applying the Hansen (1999) panel threshold model with the capacity-to-pay measure of OOP burden, we identify an empirical threshold at 11.84 percent of capacity to pay. Above this threshold, the marginal effect of OOP burden on Gini-measured inequality is approximately 17 percent greater than that observed below this threshold. A complementary structural-stability analysis reveals that the marginal inequality effect of OOP burden fell by roughly half around 2009, with no detectable additional shift at the implementation of the Affordable Care Act and no differential decline in Medicaid-expansion states.

Two broader implications follow. First, the inequality consequences of OOP financing begin to intensify at burden levels substantially below the conventional catastrophic-spending cutoffs (10, 25, and 40 percent) used in the international financial-protection literature. Headcount measures based on these cutoffs remain valuable monitoring tools, but they do not capture the distributional dynamics operating in the broader range of moderate OOP burden, i.e., a range that includes the bulk of US households during our sample period. Second, the channel is not a fixed feature of the US health-financing system: its strength has varied markedly over time, and dating that variation correctly matters for how the distributional effects of insurance reforms are assessed.

Several extensions of this work would be valuable. Decomposing the ACA effect into its component provisions (such as Medicaid expansion, marketplace subsidies, essential-benefits requirements, and the cost-sharing limit) through designs exploiting state-level variation in adoption timing would clarify which features of the reform drive recession era break extension we document. Replicating the threshold analysis on household-level data, such as the Medical Expenditure Panel Survey, would provide direct evidence on the consumption-crowding and cost-sharing channels that we identify as the most likely mechanisms underlying the regime structure. Extending the analysis to other high-income contexts with different health-financing structures would establish whether the regime dependence we document is specific to the US institutional setting or reflects a more general feature of the OOP-burden–inequality relationship. Finally, as a sufficiently long post-2019 panel becomes available, examining whether the threshold and regime structure have persisted, intensified, or attenuated in the post-pandemic period will be a natural extension of the present analysis.

Tables

Table 1. Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max	Obs.
Panel A: Outcome variable					
Gini coefficient (Frank)	0.6072	0.0350	0.5360	0.7200	816
Panel B: OOP burden measures (key explanatory variable)					
OOP burden — capacity to pay	0.1008	0.0227	0.0239	0.1710	816
OOP burden — budget share	0.0719	0.0168	0.0173	0.1245	816
Panel C: Health expenditure (\$ per capita)					
OOP health expenditure	2,502.53	661.25	1,004.00	6,126.00	816
Total health expenditure	8,823.89	1,566.56	5,470.00	14,504.00	816
Public spending (Medicare + Medicaid)	3,267.44	792.16	1,428.00	5,835.00	816
Private insurance spending	3,053.17	696.76	1,704.00	7,016.00	816
Panel D: Health financing structure (shares)					
Public spending share of THE	0.3692	0.0546	0.2456	0.5003	816
Private spending share of THE	0.3463	0.0447	0.2405	0.5114	816
Panel E: Economic and consumption variables					
Real GDP per capita (\$, chained 2017)	57,240.56	21,891.71	35,087.97	200,287.00	816
Personal consumption per capita (\$)	35,413.98	6,978.61	21,251.99	70,108.56	816
Basic-needs consumption per capita (\$)	10,144.58	1,892.64	6,405.57	19,117.60	816

Notes: Variables are constructed at the state-year level for the 51 US states (50 states plus the District of Columbia) over 2004–2019, yielding 816 state-year observations. The Gini coefficient is from the updated Frank (2009) state-level inequality series, computed on IRS-reported pre-tax adjusted gross income using the Cowell-Mehta (1982) compromise interpolation method. OOP burden under the capacity-to-pay (CTP) measure is OOP health expenditure divided by personal consumption net of basic-needs consumption (food, clothing, housing and utilities); under the budget-share measure it is OOP divided by total personal consumption. Health expenditure variables are from the IHME State-Level Health Expenditure database. Personal consumption and basic-needs consumption are from the BEA. Real GDP per capita is from the BEA in chained 2017 dollars. Population estimates used to convert aggregate measures to per capita units are from FRED.

Table 2. Diagnostic Tests on Residuals and Variables*Panel: 51 US states, 2004–2019.*

Panel A: Residual diagnostics	Statistic	p-value	Decision
Pesaran (2004) CD test (cross-sectional independence)	27.87	< 0.001	Reject H_0
Wooldridge (2010) AR(1) autocorrelation test	41.32	< 0.001	Reject H_0
Modified Wald test for groupwise heteroscedasticity	4,689.26	< 0.001	Reject H_0

Panel B: Panel unit root tests (with deterministic trend)	Fisher ADF P-stat	p-value	Pesaran CADF t-bar	p-value	Verdict
OOP burden (CTP) — threshold variable	223.07	< 0.001	−3.142	< 0.001	Stationary
Gini coefficient	147.23	0.002	−1.470	1.000	Mixed
ln(GDP per capita)	94.63	0.685	−2.175	0.771	Inconclusive
Public spending share	200.56	< 0.001	−2.049	0.949	Mixed
Private spending share	176.85	< 0.001	−1.943	0.991	Mixed

Notes: Panel A reports residual diagnostics applied to the linear two-way fixed-effects model in equation (2). The Pesaran (2004) CD test reports the standardized cross-sectional dependence statistic; the average absolute pairwise correlation across states is 0.567. The Wooldridge (2010) test reports an F-statistic with 1 and 50 degrees of freedom for the null of no first-order autocorrelation. The modified Wald test reports a χ^2 statistic with 51 degrees of freedom for the null of equal error variances across states. Panel B reports two complementary panel unit root tests, both with constant and deterministic trend. The Fisher-type augmented Dickey-Fuller test (Choi, 2001) combines panel-specific ADF tests using the inverse chi-squared statistic with 102 degrees of freedom; this test assumes cross-sectional independence. The Pesaran (2007) CADF test augments each panel's ADF regression with the cross-sectional average and is therefore robust to cross-sectional dependence; the t-bar statistic is the average of panel-specific CADF t-statistics. Both tests use the null of all panels containing unit roots. The verdict "Stationary" indicates rejection by both tests, "Inconclusive" indicates rejection by neither, and "Mixed" indicates rejection by Fisher but not CADF which is a pattern consistent with the cross-sectional dependence detected in Panel A. As discussed in Section 4.1, the OOP burden threshold variable rejects the unit root null under both tests, satisfying the stationarity requirement of the Hansen (1999) estimator.

Table 3. Baseline Linear Fixed-Effects Models*Dependent variable: Gini coefficient. Panel: 51 US states, 2004–2019*

	(1)	(2)	(3)
	Linear FE	Quadratic FE	Linear FE (DK)
OOP burden (CTP)	1.169** (0.537)	−0.282 (1.123)	1.169*** (0.357)
OOP burden (CTP) squared		5.569 (4.072)	
ln(GDP per capita)	0.005 (0.034)	0.008 (0.036)	0.005 (0.005)
Public spending share	0.613*** (0.218)	0.502** (0.224)	0.613*** (0.155)
Private spending share	0.524** (0.218)	0.409* (0.222)	0.524*** (0.137)
State fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
SE type	Cluster	Cluster	Driscoll-Kraay
F-test on squared term (p-value)		0.178	
Observations	816	816	816
Number of states	51	51	51
Within R ²	0.173	0.183	0.173

Notes: Fixed-effects (within) regressions with state and year fixed effects. OOP burden is defined under the capacity-to-pay (CTP) approach as OOP health expenditure divided by capacity to pay (personal consumption minus basic-needs consumption), following Xu et al. (2003) and Cylus et al. (2018). Column (1) reports the linear specification with cluster-robust standard errors clustered at the state level. Column (2) adds a quadratic term in OOP burden to test for smooth nonlinearity; the F-test on the squared term has p-value 0.178. Column (3) re-estimates the linear specification with Driscoll-Kraay (1998) standard errors, robust to cross-sectional dependence, serial correlation, and heteroscedasticity (maximum lag = 3). Standard errors in parentheses. Significance: *** p < 0.01, ** p < 0.05, * p < 0.10.

Table 4. Main Results: Hansen (1999) Panel Threshold Model (Capacity-to-Pay Specification)

	Coefficient	Std. Error	95% CI
Panel A: Regime-dependent coefficient on OOP burden (CTP)			
Below threshold (β_1): CTP burden < 11.84%	0.797***	(0.224)	[0.357, 1.237]
Above threshold (β_2): CTP burden $\geq$ 11.84%	0.933***	(0.218)	[0.505, 1.361]
Panel B: Threshold estimate			
$\hat{\gamma}$ (OOP / capacity to pay)	0.1184		[0.1182, 0.1185]
Panel C: Regime-independent controls			
ln(GDP per capita)	-0.001	(0.014)	[-0.029, 0.028]
Public spending share	0.520***	(0.091)	[0.341, 0.699]
Private spending share	0.467***	(0.088)	[0.294, 0.641]
Year fixed effects	Yes		
State fixed effects	Yes		
Panel D: Sequential threshold tests			
F ₁ (linear <i>versus</i> single)	34.14		p = 0.034
F ₂ (single <i>versus</i> double)	17.70		p = 0.195
Panel E: Model fit			
Observations	816		
Number of states	51		
Within R ²	0.207		

Notes: Hansen (1999) fixed-effect panel threshold model estimated via xthreg (Wang, 2015). Threshold variable and regime-dependent variable: OOP burden defined as OOP health expenditure divided by capacity to pay, where capacity to pay equals personal consumption minus basic-needs consumption (food, clothing, housing and utilities), following Xu et al. (2003) and Cylus et al. (2018). Standard errors in parentheses. 95% confidence intervals in brackets; confidence interval for the threshold constructed via inversion of the likelihood ratio statistic (Hansen, 1999). Bootstrap p-values based on 2,000 replications. The sequential test in Panel D fails to reject the null of a single threshold against a double-threshold alternative (p = 0.195), so the single-threshold model is retained as the preferred specification. All regressions include state and year fixed effects. Significance: *** p < 0.01, ** p < 0.05, * p < 0.10.

Table 5. Structural stability of the OOP burden–inequality relationship*Dependent variable: Gini coefficient. Panel: 51 US states, 2004–2019.*

	(1)	(2)	(3)	(4)	(5)	(6)
	CTP	BS	CTP	BS	CTP	CTP
<i>Break specification</i>	2014	2014	2009	2009	2009 + 2014	Expansion
OOP burden	1.464*** (0.312)	1.835*** (0.448)	1.552*** (0.260)	2.013*** (0.359)	1.600*** (0.269)	1.470*** (0.252)
OOP burden × Post-2009			−0.799*** (0.103)	−1.075*** (0.139)	−0.731*** (0.122)	−0.754*** (0.122)
OOP burden × Post-2014	−0.426** (0.181)	−0.548** (0.245)			−0.115* (0.064)	
OOP burden × Medicaid expansion						−0.084 (0.070)
Medicaid expansion active						0.015 (0.010)
ln(GDP per capita)	0.010 (0.011)	0.011 (0.012)	0.039*** (0.009)	0.044*** (0.010)	0.038*** (0.010)	0.032*** (0.007)
Public spending share	0.623*** (0.141)	0.570*** (0.138)	0.452** (0.157)	0.407** (0.153)	0.468*** (0.158)	0.377** (0.145)
Private spending share	0.561*** (0.126)	0.520*** (0.127)	0.410** (0.146)	0.385** (0.147)	0.430** (0.149)	0.404** (0.139)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	816	816	816	816	816	816
Number of states	51	51	51	51	51	51
Within R ²	0.231	0.218	0.321	0.310	0.324	0.330

Notes: The dependent variable is the state-level Gini coefficient. Columns (1) and (2) estimate a single structural break at the 2014 implementation of the ACA, with OOP burden measured under the capacity-to-pay (CTP) approach in column (1) and the budget-share (BS) approach in column (2); these columns reproduce the conventional pre/post-ACA specification. Columns (3) and (4) estimate the same single-break model with the break placed instead at 2009, the year identified by the break-date scan reported in the text, under the CTP and BS measures respectively. Column (5) includes the 2009 and 2014 break interactions jointly, so that the Post-2014 coefficient measures the additional shift at ACA implementation beyond the Recession-era break. Column (6) tests the cross-sectional implication of an ACA effect: it interacts OOP burden with an indicator for state-years with an active ACA Medicaid expansion, controlling for the 2009 break, so that a negative interaction would indicate a larger decline in expansion states. Expansion status is coded from the Kaiser Family Foundation expansion tracker, active from year t if coverage began on or before July 1 of year t and from $t + 1$ otherwise. The main effects of Post-2009 and Post-2014 are absorbed by the year fixed effects; the Medicaid expansion main effect is identified by cross-state variation in expansion timing. Driscoll-Kraay standard errors with a maximum lag of three in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6. Robustness Checks (Alternative Samples, Alternative OOP Burden Measures, and Labor-Market Controls)

Dependent variable: Gini coefficient.

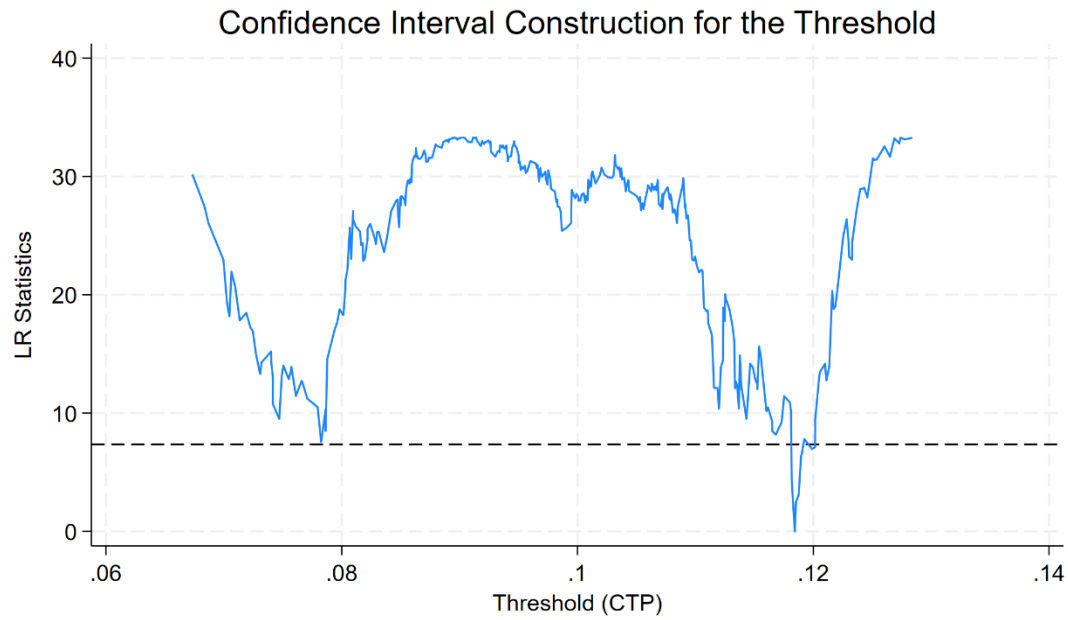
	(1)	(2)	(3)	(4)
	<i>CTP, excl. DC</i>	<i>BS, excl. DC</i>	<i>BS, double-threshold</i>	<i>CTP, unemployment control</i>
	<i>Single threshold</i>	<i>Single threshold</i>	<i>Full sample</i>	<i>Full sample</i>
Panel A: Regime-dependent coefficients				
Regime 1 (lowest burden)	0.660*** (0.230)	1.301*** (0.303)	1.046*** (0.314)	0.614*** (0.226)
Regime 2	0.793*** (0.224)	0.885*** (0.309)	0.691** (0.316)	0.753*** (0.220)
Regime 3 (highest burden)	—	—	0.841*** (0.308)	—
Panel B: Threshold estimates				
$\hat{\gamma}_1$	0.1187 [0.1183, 0.1189]	0.0533 [0.0527, 0.0536]	0.0528 [0.0522, 0.0532]	0.1184 [0.1182, 0.1185]
$\hat{\gamma}_2$	—	—	0.0835 [0.0828, 0.0837]	—
Panel C: Controls				
ln(GDP per capita)	0.005 (0.014)	0.005 (0.015)	−0.006 (0.014)	0.020 (0.015)
Public spending share	0.490*** (0.093)	0.401*** (0.094)	0.381*** (0.093)	0.460*** (0.091)
Private spending share	0.403*** (0.092)	0.255*** (0.096)	0.301*** (0.093)	0.381*** (0.090)
Unemployment rate	—	—	—	0.003*** (0.001)
State & year fixed effects	Yes	Yes	Yes	Yes
Panel D: Threshold-effect tests (bootstrap p-values)				
F ₁ (linear <i>versus</i> single)	0.027	0.034	0.078	0.025

F ₂ (single <i>versus</i> double)	—	—	0.091	—
F ₃ (double <i>versus</i> triple)	—	—	0.917	—
Panel E: Model fit				
Observations	800	800	816	816
Number of states	50	50	51	51
Within R ²	0.201	0.199	0.215	0.224

Notes: Hansen (1999) fixed-effect panel threshold model estimated via xthreg (Wang, 2015). Column (1) re-estimates the main CTP single-threshold specification excluding the District of Columbia. Column (2) re-estimates the budget-share single-threshold specification excluding the District of Columbia. Column (3) reports the full-sample double-threshold specification under the budget-share measure. Column (4) re-estimates the main CTP single-threshold specification on the full sample with the state unemployment rate (BLS Local Area Unemployment Statistics, annual average) added to the control set. CTP = capacity-to-pay measure (OOP / (personal consumption – basic needs)). BS = budget-share measure (OOP / personal consumption). Standard errors in parentheses. 95% confidence intervals in brackets. Bootstrap p-values based on 1,000 replications for the exclude-DC and unemployment-control specifications and 2,000 replications for the full-sample double-threshold specification (300 replications for the triple-threshold test). All regressions include state and year fixed effects. Significance: *** p < 0.01, ** p < 0.05, * p < 0.10.

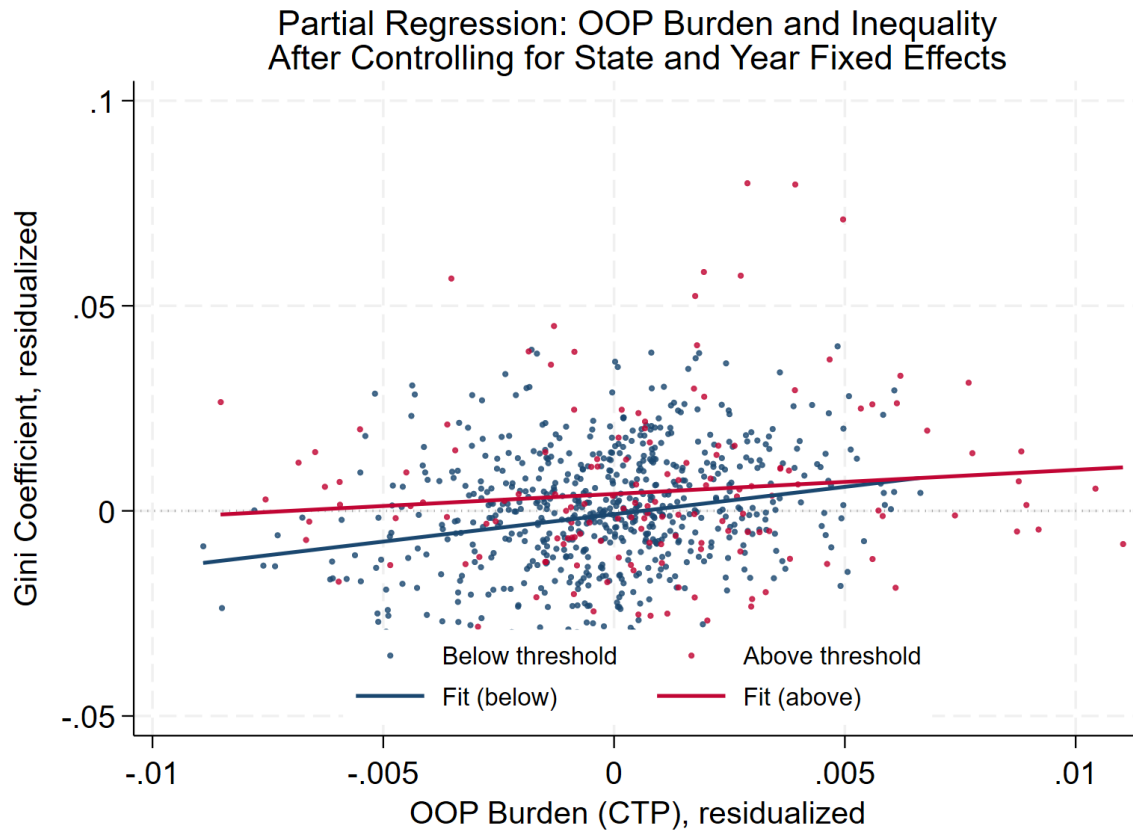
Figures

Figure 1. Confidence Interval Construction for the Threshold Estimates



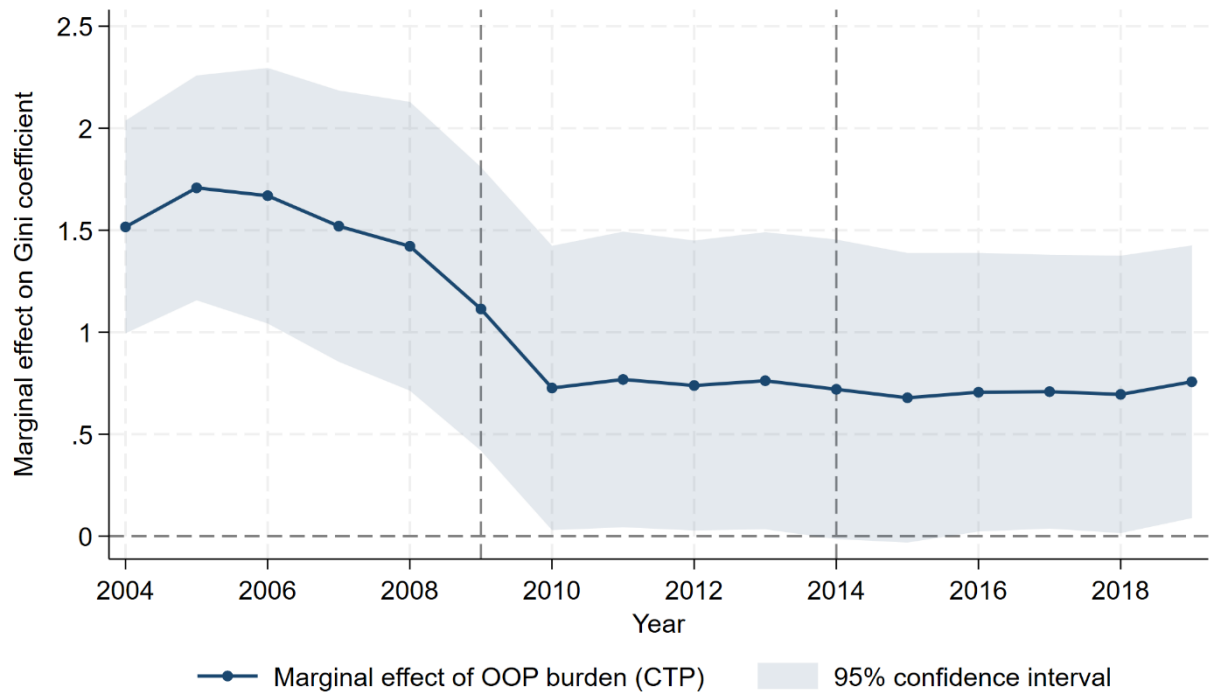
Notes: Likelihood ratio (LR) statistics are plotted against candidate threshold values for the single-threshold Hansen (1999) model under the capacity-to-pay (CTP) measure. The dashed horizontal line denotes the 95% critical value (7.35); the 95% confidence interval is the set of values for which the LR statistic lies below the critical value. The profile is V-shaped and centered at $\hat{\gamma} = 0.1184$ (95% CI [0.1182, 0.1185]), corresponding to an OOP burden of 11.84% of capacity to pay.

Figure 2. Within-State Effect of OOP Burden on Inequality, by Regime



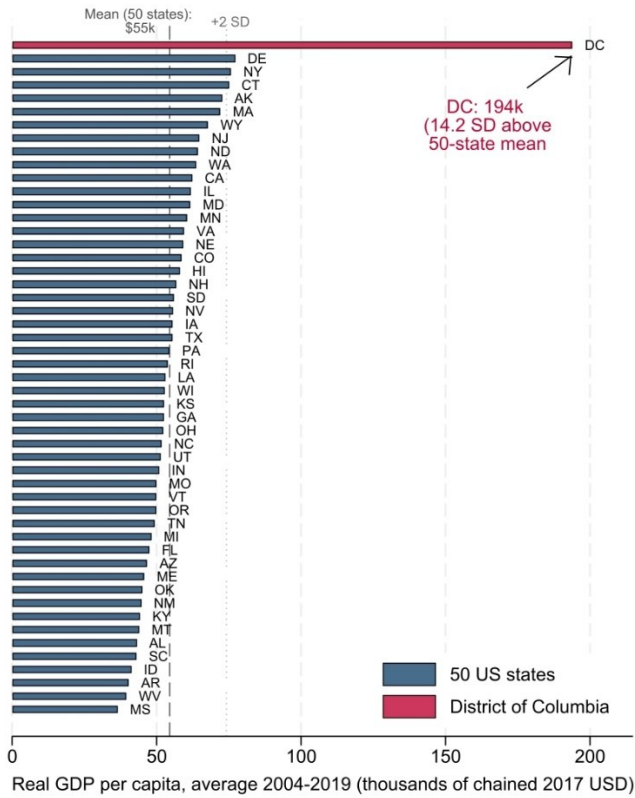
Notes: Partial regression plot of state-level Gini coefficient against OOP burden (capacity-to-pay measure), with both variables residualized on state fixed effects, year fixed effects, log GDP per capita, public spending share, and private spending share. Each point represents one state-year. Vertical position reflects the unexplained variation in the Gini after controlling for state, year, and macroeconomic factors. Linear fit lines are estimated separately for state-years with OOP burden below and above the estimated threshold ($\hat{\gamma} = 0.1184$).

Figure 3. Year-by-year marginal effect of OOP burden (CTP) on the Gini coefficient



Notes: Point estimates and 95% confidence intervals for the marginal effect of OOP burden in each year, from the fixed-effects specification augmented with interactions between OOP burden and the full set of year indicators; 2004 is the base year. Driscoll-Kraay standard errors with a maximum lag of three. Dashed vertical lines mark 2009 and 2014.

Figure 4. The District of Columbia as an outlier on real GDP per capita, 2004–2019



Notes: Bars show average real GDP per capita, 2004–2019, in thousands of chained 2017 dollars. Dashed line: mean across the 50 states (DC excluded); dotted line: +2 SD. DC sits 14 SDs above the 50-state mean. Sources: BEA, FRED, authors' calculations.

References

- Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61(4), 821. <https://doi.org/10.2307/2951764>
- Ashby, N. J., & Sobel, R. S. (2007). Income inequality and economic freedom in the U.S. states. *Public Choice* 2007 134:3, 134(3), 329–346. <https://doi.org/10.1007/S11127-007-9230-5>
- Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3), 244–263. [https://doi.org/10.1016/0022-0531\(70\)90039-6](https://doi.org/10.1016/0022-0531(70)90039-6)
- Buettgens, M., Blavin, F., & Pan, C. (2021). The Affordable Care Act Reduced Income Inequality In The US. *Health Affairs*, 40(1), 121–129. <https://doi.org/10.1377/HLTHAFF.2019.00931>
- Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20(2), 249–272. [https://doi.org/10.1016/S0261-5606\(00\)00048-6](https://doi.org/10.1016/S0261-5606(00)00048-6)
- Christopher, A. S., Himmelstein, D. U., Woolhandler, S., & McCormick, D. (2018). The effects of household medical expenditures on income inequality in the United States. *American Journal of Public Health*, 108(3), 351–354. <https://doi.org/10.2105/AJPH.2017.304213>
- Courtemanche, C., Marton, J., Ukert, B., Yelowitz, A., Zapata, D., & Fazlul, I. (2019). The three-year impact of the Affordable Care Act on disparities in insurance coverage. *Health Services Research*, 54 Suppl 1(Suppl 1), 307–316. <https://doi.org/10.1111/1475-6773.13077>
- Cowell, F. A., & Mehta, F. (1982). The Estimation and Interpolation of Inequality Measures. *The Review of Economic Studies*, 49(2), 273–290. <https://doi.org/10.2307/2297275>
- Cylus, J., Thomson, S., & Evetovits, T. (2018). Catastrophic health spending in Europe: equity and policy implications of different calculation methods. *Bulletin of the World Health Organization*, 96(9), 599–609. <https://doi.org/10.2471/BLT.18.209031>
- De Maio, F. G. (2007). Income inequality measures. *Journal of Epidemiology & Community Health*, 61(10), 849–852. <https://doi.org/10.1136/JECH.2006.052969>
- Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data. *The Review of Economics and Statistics*, 80(4), 549–560. <https://doi.org/10.1162/003465398557825>
- Frank, M. W. (2009). Inequality and growth in the United States: Evidence from a new state-level panel of income inequality measures. *Economic Inquiry*, 47(1), 55–68. <https://doi.org/10.1111/J.1465-7295.2008.00122.X>

- Galbraith, A. A., Ross-Degnan, D., Soumerai, S. B., Rosenthal, M. B., Gay, C., & Lieu, T. A. (2011). Nearly Half of Families in High-Deductible Health Plans Whose Members Have Chronic Conditions Face Substantial Financial Burden. *Health Affairs*, 30(2), 322–331. <https://doi.org/10.1377/HLTHAFF.2010.0584>
- Goldman, A. L., Woolhandler, S., Himmelstein, D. U., Bor, D. H., & McCormick, D. (2018). Out-of-Pocket Spending and Premium Contributions After Implementation of the Affordable Care Act. *JAMA Internal Medicine*, 178(3), 347–355. <https://doi.org/10.1001/JAMAINTERNMED.2017.8060>
- Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93(2), 345–368. [https://doi.org/10.1016/S0304-4076\(99\)00025-1](https://doi.org/10.1016/S0304-4076(99)00025-1)
- Hatch, M. E., & Rigby, E. (2015). Laboratories of (In)equality? Redistributive Policy and Income Inequality in the American States. *Policy Studies Journal*, 43(2), 163–187. <https://doi.org/10.1111/PSJ.12094>
- Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal*, 7(3), 281–312. <https://ideas.repec.org/a/tsj/stataj/v7y2007i3p281-312.html>
- Kaiser Family Foundation. (2019). *2019 Employer Health Benefits Survey*. <https://www.kff.org/health-costs/report/2019-employer-health-benefits-survey/>
- Ketsche, P., Adams, E. K., Wallace, S., Kannan, V. D., & Kannan, H. (2011). Lower-income families pay A higher share of income toward national health care spending than higher-income families do. *Health Affairs*, 30(9), 1637–1646. <https://doi.org/10.1377/hlthaff.2010.0712>
- Liu, C., Gotanda, H., Khullar, D., Rice, T., & Tsugawa, Y. (2021). The Affordable Care Act's Insurance Marketplace Subsidies Were Associated with Reduced Financial Burden For US Adults. *Health Affairs*, 40(3), 496–504. <https://doi.org/10.1377/HLTHAFF.2020.01106>
- Nguyen, H. A., Ahmed, S., & Turner, H. C. (2023). Overview of the main methods used for estimating catastrophic health expenditure. *Cost Effectiveness and Resource Allocation* 2023 21:1, 21(1), 50-. <https://doi.org/10.1186/s12962-023-00457-5>
- Oyekola, O. (2021). Finance and inequality in a panel of US States. *Empirical Economics*, 61(5), 2739–2795. <https://doi.org/10.1007/S00181-020-01976-3>
- Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2), 265–312. <https://doi.org/10.1002/JAE.951>

Piketty, T., & Saez, E. (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, 118(1), 1–41. <https://doi.org/10.1162/00335530360535135>

Piketty, T., Saez, E., & Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. *The Quarterly Journal of Economics*, 133(2), 553–609. <https://doi.org/10.1093/QJE/QJX043>

Sirag, A., & Nor, N. M. (2021). Out-of-pocket health expenditure and poverty: Evidence from a dynamic panel threshold analysis. *Healthcare*, 9(5). <https://doi.org/10.3390/healthcare9050536>

Smith, V. K. (2021). A note on income inequality at the state level. *Economics Letters*, 203, 109865. <https://doi.org/10.1016/j.econlet.2021.109865>

Wagstaff, A., Flores, G., Hsu, J., Smitz, M. F., Chepynoga, K., Buisman, L. R., van Wilgenburg, K., & Eozenou, P. (2018). Progress on catastrophic health spending in 133 countries: a retrospective observational study. *The Lancet Global Health*, 6(2), e169–e179. [https://doi.org/10.1016/S2214-109X\(17\)30429-1](https://doi.org/10.1016/S2214-109X(17)30429-1)

Wagstaff, A., & van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: With applications to Vietnam 1993–1998. *Health Economics*, 12(11), 921–933. <https://doi.org/10.1002/hec.776>

Wang, Q. (2015). Fixed-effect panel threshold model using Stata. *Stata Journal*, 15(1), 121–134. <https://ideas.repec.org/a/tsj/stataj/v15y2015i1p121-134.html>

Wooldridge, J. M. . (2010). *Econometric analysis of cross section and panel data*. 1064.

Xu, K., Evans, D. B., Zeramdini, R., Klavus, J., & Murray, C. J. (2003). Household catastrophic health expenditure: a multicountry analysis. *The Lancet*, 362(362), 111–117. [https://doi.org/https://doi.org/10.1016/S0140-6736\(03\)13861-5](https://doi.org/https://doi.org/10.1016/S0140-6736(03)13861-5)

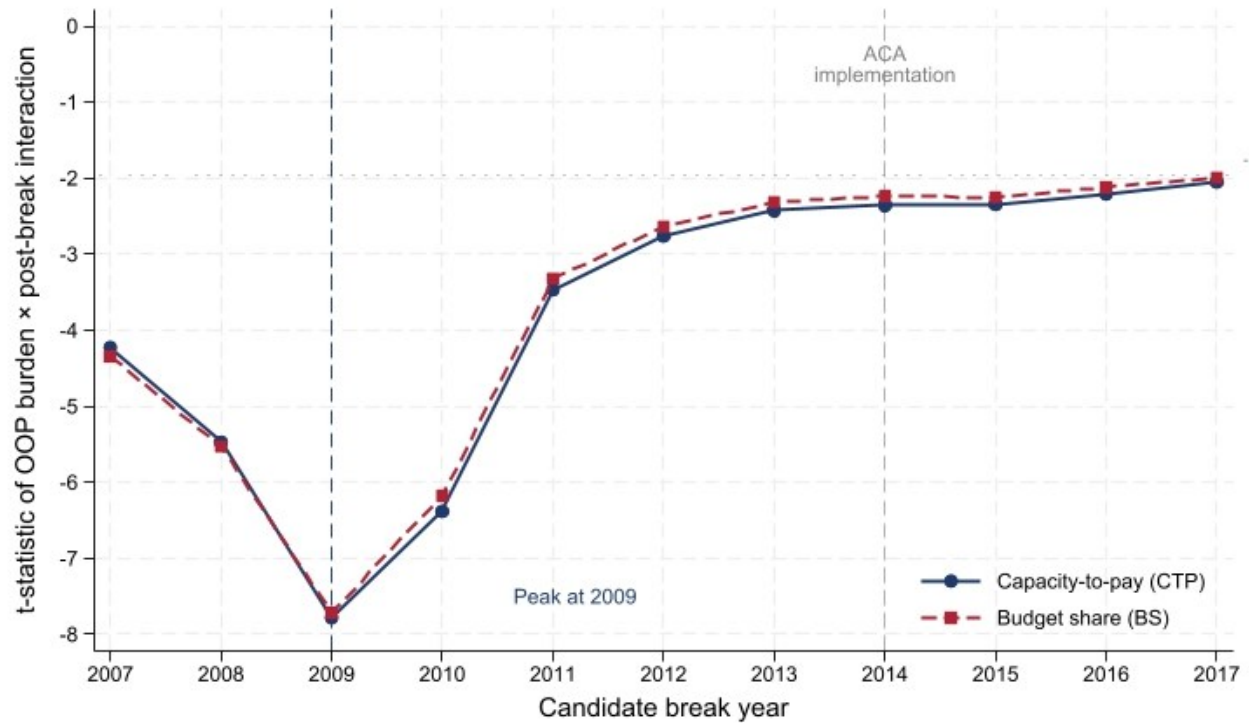
Appendix A

Table A1. Panel Variation Decomposition

Variable		Mean	Std. Dev.	Min	Max	Obs.
<i>Gini coefficient</i>	overall	0.6072	0.0350	0.5360	0.7200	N = 816
	between		0.0300	0.5607	0.6868	n = 51
	within		0.0185	0.5573	0.6934	T = 16
<i>OOP burden (CTP)</i>	overall	0.1008	0.0227	0.0239	0.1710	N = 816
	between		0.0207	0.0359	0.1535	n = 51
	within		0.0099	0.0697	0.1298	T = 16
<i>OOP burden (budget share)</i>	overall	0.0719	0.0168	0.0173	0.1245	N = 816
	between		0.0155	0.0263	0.1118	n = 51
	within		0.0069	0.0481	0.0919	T = 16
<i>OOP per capita (\$)</i>	overall	2,502.53	661.25	1,004.00	6,126.00	N = 816
	between		607.26	1,540.38	4,710.19	n = 51
	within		274.37	709.35	3,918.35	T = 16
<i>Total health exp. per capita</i>	overall	8,823.89	1,566.56	5,470.00	14,504.00	N = 816
	between		1,295.05	6,325.63	12,901.00	n = 51
	within		898.80	6,054.02	11,866.02	T = 16
<i>Real GDP per capita</i>	overall	57,240.56	21,891.71	35,087.97	200,287.00	N = 816
	between		21,808.08	36,393.40	193,870.70	n = 51
	within		3,522.49	37,394.62	74,458.17	T = 16
<i>Public share of THE</i>	overall	0.3692	0.0546	0.2456	0.5003	N = 816
	between		0.0515	0.2635	0.4833	n = 51
	within		0.0193	0.3091	0.4273	T = 16
<i>Private share of THE</i>	overall	0.3463	0.0447	0.2405	0.5114	N = 816
	between		0.0412	0.2664	0.4803	n = 51
	within		0.0183	0.2988	0.4291	T = 16

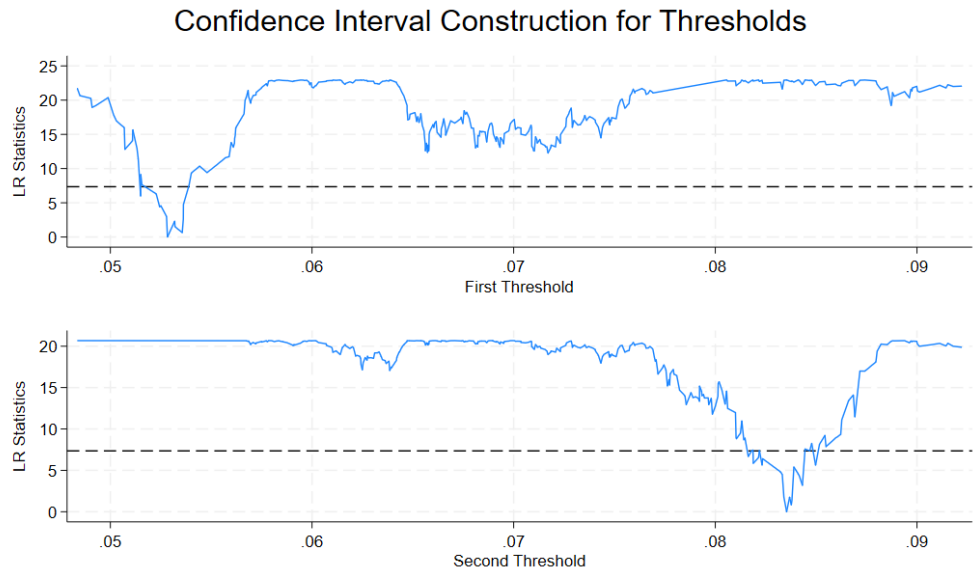
Notes: For each variable, the table reports the overall (pooled), between-state (cross-sectional), and within-state (over time) standard deviations. The within transformation removes state means; the between transformation collapses the data to state-level averages. Min and max for the within rows refer to the deviation from state means added back to the global mean. The within standard deviations are particularly relevant for fixed-effects identification: the threshold variable (*OOP burden CTP*) has within standard deviation of 0.0099, with within-state values ranging from 0.0697 to 0.1298, providing sufficient within-state variation to identify regime-switching across the estimated threshold of 0.1184 reported in Section 4.3.

Figure A1. Break-date scan for the OOP burden–inequality relationship



Note: The figure plots the Driscoll-Kraay t-statistic of the OOP burden × post-break interaction from single-break specifications estimated at each candidate break year from 2007 to 2017, under the capacity-to-pay measure (navy, solid) and the budget-share measure (crimson, dashed). Each specification includes state and year fixed effects and the full control set; the post-break main effect is absorbed by the year fixed effects. The dotted horizontal line marks the conventional two-sided 5% critical value. Under both measures the test statistic peaks sharply in 2009 and declines monotonically thereafter, with no secondary peak at the 2014 implementation of the ACA (gray vertical line), indicating a single dominant break located at the Great Recession.

Figure A2. Confidence Interval Construction for the Budget-Share Double-Threshold Model



Notes: LR statistics for the first (top panel) and second (bottom panel) thresholds of the double-threshold model under the budget-share measure. The dashed line denotes the 95% critical value (7.35). Thresholds are estimated at $\hat{\gamma}_1 = 0.0528$ (95% CI [0.0522, 0.0532]) and $\hat{\gamma}_2 = 0.0835$ (95% CI [0.0828, 0.0837]).

