An Anti-IV Approach for Pricing Residential Amenities: Applications to Flood Risk*

Alex Bell † Stephen Billings‡ Sophie Calder-Wang§ Shusheng Zhong¶

March 24, 2026

Abstract

Traditional hedonic methods have often struggled to price residential amenities due to unobserved housing quality, resulting in wrong-signed estimates. This paper introduces a novel "anti-instrument" approach to estimating amenity prices. Identification relies on an anti-instrumental variable that is relevant for unobserved quality but, conditional on that quality, uncorrelated with the focal amenity. We micro-found reported income at the time of mortgage application as an anti-IV to estimate the implicit price of flood risk. Using detailed U.S. housing transactions, mortgage records, and house-level flood risk measures, we find larger discounts for homes with higher flood probabilities, with the riskiest category suffering a discount of 3.5%. Our approach delivers both the correct sign and an economically meaningful magnitude for the implicit price of flood risk, which helps address the mixed evidence in the existing literature.

Keywords: Hedonic Pricing, Unobserved Quality, Amenities, Flood Risk
JEL Codes: R21, R32, Q51, Q54, C21, C14

*We have benefited from conversations with conference and seminar participants at AREUEA, SEA, UEA, NBER Summer Institute (Real Estate and Urban Economics), NYU Summer Climate Finance Conference, University of Toronto, SUNY Albany, and University of Georgia. This paper supersedes the methods section of a previous working paper titled "Pricing Neighborhood Amenities: A Proxy-Based Approach" with Alex Bell, Sophie Calder-Wang, and Shusheng Zhong (2023). We thank Joshua Boyd for excellent research assistance. Calder-Wang thanks the Zell/Lurie Real Estate Center at Wharton for financial support. Sample data and code to implement the methods proposed in the paper is available as a Stata package *aivreg* on GitHub.

†Department of Economics, Georgia State University *ambell@gsu.edu*
‡Leeds School of Business, University of Colorado Boulder *stephen.billings@colorado.edu*
§The Wharton School, University of Pennsylvania *sophiecw@wharton.upenn.edu*
¶Kellogg School of Management, Northwestern University *shusheng.zhong@kellogg.northwestern.edu*

1 Introduction

Understanding the prices of residential amenities is fundamental to both urban economics and environmental policy. Housing can be thought of as a bundle of amenities: beyond the usual structural characteristics such as the number of bedrooms, bathrooms, building age, square footage, lot size, etc., we are also concerned with the valuation of local amenities, including crime, congestion, school quality, and environmental disamenities such as air pollution, fire, or flood risk. The implicit prices of these residential amenities, defined as the marginal change in home price due to a marginal change in the amenity while holding all other attributes fixed, directly affect how private investment decisions are made, and are crucial inputs for assessing the returns to mitigating environmental externalities.

Unfortunately, empirical estimates based on observational studies are often unsatisfactory in providing credible estimates of amenity prices. These estimates frequently suffer from a "wrong-sign" problem due to the pervasive endogenous correlation between observed amenities and unobserved housing quality (Chay and Greenstone, 2005). In Figure 2 panel (a), for example, we show that a simple regression of home prices on flood risk, even with granular controls for housing characteristics and geography, typically yields a flood "premium," contradicting both theory and intuition. Given the importance of correctly pricing flood risk, this wrong sign and unknown magnitude problem makes it challenging to design efficient policy solutions. This problem can be especially pronounced in the housing context, likely because "all houses are, by nature, a unique combination of aspects of location and structure" (Giglio, Kelly, and Stroebel, 2021). As a result, researchers devote significant effort to finding credible research designs through quasi-experiments to estimate how changes in amenities are capitalized into housing prices (Kuminoff, Parmeter, and Pope, 2010; Bishop, Kuminoff, Banzhaf, Boyle, Von Gravenitz, Pope, Smith, and Timmins, 2020).

In this paper, we make two main contributions. First, we develop a novel *anti-instrumental variable* approach to identify and estimate implicit prices in hedonic models of housing markets in the presence of unobservable quality as a key confounder, without relying on conventional quasi-experiments. Second, we provide new estimates of the implicit price of flood risk for the US housing market. The importance of obtaining the correct sign and magnitude of the price of flood risk is of first-order policy relevance as we attempt to mitigate the increased risk of floods and other natural disasters under even conservative estimates of climate change (Calvin, Dasgupta, Krinner, Mukherji, Thorne, Trisos, Romero, Aldunce, Barrett, Blanco, Cheung, Connors, Denton, Diongue-Niang, Dodman, Garschagen, Geden, Hayward, Jones, Jotzo, Krug, Lasco, Lee, Masson-Delmotte, Meinshausen, Mintenbeck, Mokssit, Otto, Pathak, Pirani, Poloczanska, Pörtner, Revi, Roberts, Roy, Ruane, Skea, Shukla, Slade, Slangen, Sokona, Sörensson, Tignor, Van Vuuren, Wei, Winkler, Zhai, Zommers, Hourcade, Johnson, Pachauri,

Simpson, Singh, Thomas, Totin, Arias, Bustam, 2023).

We term the variable an "anti-instrument" to reflect its opposite econometric property relative to conventional instruments: it is *correlated* with the unobservable component (e.g., the unobserved quality of a home) but *uncorrelated* with the endogenous variable (e.g., flood risk), conditional on that unobserved component. Unlike methods that rely on quasi-random variations in the assignment of amenities that are uncorrelated with unobserved quality, the anti-IV approach shifts our attention to finding a single anti-IV (e.g., reported income at the time of mortgage application) that acts as a "shifter" for unobserved quality but is conditionally uncorrelated with the focal amenity. Intuitively, given that reported income may be a relevant but imperfect measure of unobserved housing quality, conditional orthogonality requires that how reported income *mismeasures* unobserved housing quality be uncorrelated with flood risk. In this way, the anti-IV approach draws on a fundamentally different source of identifying variation than conventional quasi-experimental methods.

We begin with a linear hedonic model of price, determined by observable amenities as well as unobserved quality, which can be thought of as representing a combination of all hard-to-observe qualities of a house, such as curb appeal, build quality, and maintenance. Under this linear setting, we require the candidate anti-instrument to satisfy two key econometric assumptions (i) *relevance* and (ii) *conditional orthogonality*. The relevance condition requires that the anti-IV be correlated with the unobserved quality. The conditional orthogonality condition requires the anti-instrumental variable to be *uncorrelated* with the focal amenity after conditioning on unobserved quality. Importantly, conditional orthogonality does not imply total orthogonality between the anti-instrument and the focal amenity, nor does total orthogonality imply conditional orthogonality.

With these two assumptions, we show that the implicit price can be identified as the ratio of two linear projection coefficients, which naturally produces a ratio-of-coefficient estimator. We provide its associated econometric properties, including consistency, asymptotic normality, and weak-IV robust inference. We also generalize to 2SLS and GMM estimators when the model is over-identified. These properties of the anti-IV estimator can be viewed as the parallel properties of the conventional IV estimators. It is worth clarifying that we do not seek to advance econometric technique here; rather, we use standard tools to construct estimators that implement this new identification argument. The resulting estimator is simple and computationally cheap, and is made available as a Stata package *aivreg*.

Then, we generalize our anti-IV approach to the non-parametric setting, where we no longer impose any functional form restrictions on the hedonic price function, allowing the unobserved component to enter non-separably and non-additively. The only assumption we retain is a single-index assumption, meaning that the hedonic pricing function is strictly increasing in a single index of unobserved quality. This assumption is automatically satisfied by construction

under the linear setting; in non-parametric settings, Bajari and Benkard (2005) show that the single-index assumption is minimally restrictive among hedonic models. Further, existing work to address unobservables in housing markets, such as Bajari, Fruehwirth, Kim, and Timmins (2012) and Diamond and Diamond (2024), have also relied on such single index assumptions, albeit with a different empirical strategy leveraging panel data.[1]

Analogous to the linear setting, non-parametric identification follows when the anti-IV variable is relevant, at least locally, and conditionally independent of the focal amenity. As such, the anti-IV approach can recover amenity prices that vary flexibly across different levels of price, amenity, or quality segments within the market, a desirable feature for hedonic models in general (Kuminoff, Parmeter, and Pope, 2010).

Next, we employ the anti-IV approach to estimate the price of flood risk, an economically important question given the substantial costs associated with flooding events. We assemble a nationwide dataset of U.S. single-family home transactions from 2018 to 2023 from local assessor records, combined with property-level flood risk measures from First Street, as well as FEMA (Federal Emergency Management Agency) maps. We further merge the housing transactions with income and demographics of the mortgage borrower from HMDA (the Home Mortgage Disclosure Act), geological features such as distance-to-coast and elevation from US Geological Survey, and, where available, party affiliation of registered voters from L2.

To estimate the price of flood risk, we propose reported income at the time of mortgage application as the candidate anti-IV, which we further microfound through a model of housing choice. The key insight is that household characteristics, such as income, can be particularly informative about the unobserved quality of the chosen house. Moreover, within a class of housing choice models where consumption of unobserved quality expands directly with household budget,[2] we can interpret conditional orthogonality as requiring that the extent to which reported income mismeasures household budget must be uncorrelated with their flood preferences. Thus, even though reported income is not typically considered a "hedonic" characteristic of the house and is not a structural input into the pricing function, it can serve as an anti-instrument to recover amenity prices in the presence of unobserved quality.

Using reported income as the anti-IV, we find that properties at the greatest risk of flooding (First Street Foundation Flood Factor 9 and 10) sell at *discounts* of 0.7% and 3.5% , respectively. This stands in sharp contrast with conventional OLS regressions, which produce intuitively wrong-signed estimates even after extensively controlling for housing attributes and granular census block fixed effects. Relatedly, applying the anti-IV approach to FEMA's 100-year flood-

[1] We do not focus on panel methods because our setting does not involve time-varying amenities.

[2] It is common in discrete choice models (Berry, Levinsohn, and Pakes, 1995; Miravete, Seim, and Thurk, 2024), including housing choice models (Bayer, Ferreira, and McMillan, 2007; Galiani, Murphy, and Pantano, 2015; Calder-Wang, 2021; Cook, Li, and Binder, 2024), to allow price sensitivity to vary by household budget.

plain designation yields a 2.9% discount, an amount broadly comparable to the present value of lender-required flood insurance costs. Furthermore, guided by the housing choice model, we assess anti-IV validity by adding controls that may co-vary with preferences for flood risk or with the mismeasurement of budget by reported income. Adding controls for party affiliation, race and ethnicity, and dummies for age cohorts does not materially affect the estimates, strengthening our confidence in the main estimates.

Our empirical application shows that the anti-IV approach not only generates estimates of flood risk that are correctly signed but also yields magnitudes that increase with underlying severity. The ability to quantify the magnitude of flood risk is especially relevant, both because of its direct policy implications and because recent studies of sea level rise risk have reported mixed findings (Bernstein, Gustafson, and Lewis, 2019; Murfin and Spiegel, 2020; Baldauf, Garlappi, and Yannelis, 2020). Variation in conclusions from this literature appears to be due to different research designs rather than datasets, likely reflecting the inherent difficulty of fully controlling for unobserved confounders. In addition, scholars have estimated flood risk using home price changes following flood events (Bin and Landry, 2013; Ortega and Taşpınar, 2018; Beltrán, Maddison, and Elliott, 2019; Gibson and Mullins, 2020). These studies require an additional assumption, namely that property-level attributes or non-focal amenities remain unchanged by the event, which is not required by our approach.

Given the extensive literature on hedonic models following Rosen (1974), we emphasize that the goal of the anti-IV approach is to provide a new identification strategy for the hedonic price function in the presence of unobserved quality as the main confounder. Given that recovering implicit prices has been the dominant focus of the applied hedonic literature, we believe it is particularly valuable to develop new paths towards identification beyond quasi-experiments. Our paper also naturally complements the subsequent step to further identify underlying primitives, namely, demand preferences or production technologies, where a large literature has developed,[3] and for which an appropriately estimated hedonic price surface is needed as the input. Furthermore, by not requiring quasi-random changes in amenities, the anti-IV approach can be applied within a single cross-section and identifies the implicit price at a given market equilibrium. This property avoids complications arising from endogenous resorting after amenity changes (Kuminoff and Pope, 2014).

Lastly, our paper also contributes to several strands of methods literature where we show how conditional orthogonality (or conditional independence) can be used to deliver identification.[4]

[3] A select subset of this literature includes Brown and Rosen (1982); Epple (1987); Ekeland, Heckman, and Nesheim (2004); Heckman, Matzkin, and Nesheim (2010); Bajari and Benkard (2005); Bishop and Timmins (2019); Chernozhukov, Galichon, Henry, and Pass (2021).

[4] We discuss in greater detail later how our paper extends and broadens methodological innovations in the selection literature (Altonji, Elder, and Taber, 2005), production estimation (Olley and Pakes, 1996), and the causal inference literature (Miao, Geng, and Tchetgen Tchetgen, 2018; Deaner, 2023).

Although our paper centers on the hedonic model of housing, our identification argument can be applied in other settings if the single index assumption is appropriate.[5] Similar to using conventional IV strategies, the main empirical challenge remains finding an anti-instrument that plausibly satisfies both relevance and conditional orthogonality. To that end, our empirical application highlights that a combination of economic relationships and institutional knowledge can help generate suitable candidates for anti-IVs.

The remainder of the paper is organized as follows. Section 2 introduces the anti-IV approach as a method for identifying and estimating implicit prices in hedonic models of housing. Section 2.2 begins with a linear price function. It presents the main identification results and derives closed-form solutions for linear estimators. Section 2.3 generalizes it to the non-parametric setting. Section 3 describes the main dataset of housing transactions. Section 4 examines the applicability of the key assumptions in pricing flood risk, including a model of housing choice used to micro-found reported income as an anti-IV. Section 5 presents the empirical results. Section 6 concludes.

2 The Anti-IV Approach in Pricing Housing Amenities

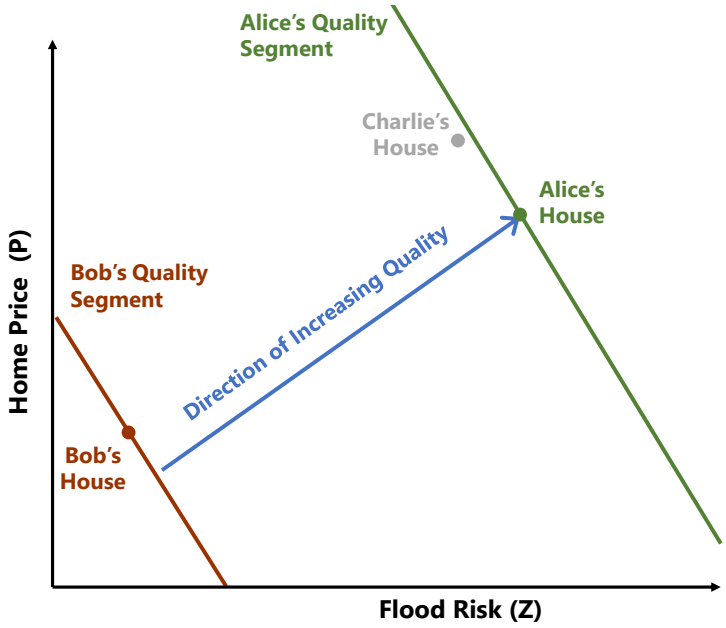
2.1 Motivation

To motivate our method, we begin with an example that highlights the role of unobservable quality in estimating amenity prices. Consider Alice and Bob. Alice lives in a nice, expensive, oceanfront single-family home with high flood risk. Bob lives in a similarly sized but less expensive single-family home in the same neighborhood, farther away from the shore and at lower flood risk. A naive regression of price on flood risk would wrongly imply a positive price for flood risk (i.e., homes with more flood risk are more expensive). The issue here is a form of omitted-variable bias: Alice's home is likely better than Bob's in many ways not controlled by this naive regression, such as beach access, nicer interiors, or better maintenance. Figure 1 illustrates this idea by placing Alice's house on a higher segment of housing quality than Bob's.

To correct this, we need to compare houses with different flood risks that otherwise sit in the same quality segment. For example, Charlie's house might be a good candidate: it is located on the same block and also has an oceanfront view, but is less susceptible to flood risk because the shoreline curves at a different angle. Conventional hedonic regression approaches would attempt to address the omitted quality by including as many observable controls for quality as possible.

[5] Our paper provides a conceptual framework to formalize the assumptions needed to use such approaches developed in the labor literature where an imperfect measure of the confounder is used as the outcome variable in a regression, which include early studies of salary discrimination (Birnbaum, 1979a,b; Hashimoto and Kochin, 1980; Kapsalis, 1982; Conway and Roberts, 1983; Goldberger, 1984) and more recent studies of compensating wage differentials (Bell, 2020; Folke and Rickne, 2022; Burbano, Folke, Meier, and Rickne, 2022; Maideu-Morera, 2024; De Schouwer and Kesternich, 2024; Ferreira, de Morentin, and Erro-Garcés, 2025).

Figure 1: Motivation for the Anti-IV Approach



However, a fundamental issue in housing is that every house is inherently unique in its location and characteristics, making it infeasible to control for all features that contribute to its price. If the remaining unobservable features remain correlated with the amenity, they continue to bias the estimate. In this example, it could be that Charlie's house, comparable to Alice's in many observable ways, might face a less appealing, rocky stretch of beach and has less well-maintained interiors. Thus, Charlie's house, while lower in flood risk, still lies on a lower quality segment than Alice's, contributing to the bias.

The core innovation of our paper is that we can eschew the need to control for *all* price-relevant housing features by using just *one* anti-instrumental variable that is relevant for unobserved quality and satisfies an exclusion restriction of conditional orthogonality. In essence, this anti-instrument proxies for a composite measure of unobserved quality; then, by conditioning on the component of the anti-IV that covaries with unobserved quality, we can recover a consistent estimate of the amenity price.

2.2 Identification in the Linear Model

To begin, we derive our main identification results and estimation procedure assuming a linear functional form in the hedonic price function. Focusing on a linear functional form mimics how applied researchers overwhelmingly estimate hedonic price functions and provides the simplest framework to introduce our anti-instrumental variable approach. Subsequently, we provide a full generalization to achieve non-parametric identification in the next subsection.

2.2.1 Assumptions

Assumption 2.1. ***(Linearity)*** The hedonic price function $P(Z, \Phi)$ exists and is linear in observed amenity Z and unobserved quality Φ, namely,

$$P = \beta_0 + \beta_z Z + \Phi. \tag{2.1}$$

The linearity of the hedonic price function is an economic assumption about the underlying structure of how prices are determined. The key object of interest is the true implicit price of the amenity, β_z. Since the unobserved quality, Φ, is a scale-free parameter, we normalize its coefficient to 1. Importantly, the linear setting guarantees that the unobserved quality can be inverted and is also a linear function of price and amenity, namely,

$$\Phi(P, Z) = P - \beta_0 - \beta_z Z. \tag{2.2}$$

Next, we introduce a candidate anti-instrumental variable H. We define *linear predicted anti-IV* $\tilde{H}(P, Z)$ as the linear projection of H on price P and amenity Z. Correspondingly, we define $\tilde{H}(\Phi)$ as its linear projection on unobserved quality Φ.

Definition 2.2. (Linear Predicted Anti-IV)

$$\tilde{H}(P, Z) \equiv \mathbb{E}^*[H|P, Z] = \delta_0 + \delta_p P + \delta_z Z \tag{2.3}$$

$$\tilde{H}(\Phi) \equiv \mathbb{E}^*[H|\Phi] \quad = \gamma_0 + \gamma_\phi \Phi \tag{2.4}$$

where $\mathbb{E}^*[\cdot]$ is the linear projection operator. $(\delta_0, \delta_p, \delta_z)$ and (γ_0, γ_ϕ) denote the respective linear projection coefficients.

Now, we state the main econometric assumptions for an anti-IV H.

Assumption 2.3. ***(Relevance)***

$$\text{cov}(H, \Phi) \neq 0. \tag{2.5}$$

The relevance assumption stipulates that the anti-instrumental variable H is, on average, informative about the quality index Φ. Importantly, since $\tilde{H}(\Phi) = \mathbb{E}^*[H|\Phi] = \gamma_0 + \gamma_\phi \Phi$, it is a linear projection of H on Φ, so the relevance condition guarantees that $\gamma_\phi \neq 0$. In other words, the relevance condition guarantees that there is a linear one-to-one mapping between the linear predicted IV $\tilde{H}(\Phi)$ and unobserved quality Φ.

Assumption 2.4. ***(Conditional Orthogonality)*** An anti-instrumental variable H is conditionally orthogonal to Z if

$$\text{cov}(H^{\perp\Phi}, Z) = 0, \tag{2.6}$$

where $H^{\perp\Phi}$ denotes the residual from the linear projection of H on Φ.

In other words, a variable H satisfies *conditional* orthogonality if the remaining variation in H after we condition on unobserved quality Φ is uncorrelated with amenity Z. Equivalently, conditional orthogonality also implies that the linear projection coefficient of H on Z is zero once we project on Φ, namely,

$$\mathbb{E}^*[H|\Phi, Z] = \mathbb{E}^*[H|\Phi]. \tag{2.7}$$

It is worth noting that conditional orthogonality does *not* imply unconditional orthogonality (i.e., $\text{cov}(H, Z) = 0$), and neither does unconditional orthogonality imply conditional orthogonality. Also, notice that Eq (2.6) represents a form of conditional orthogonality in which the unobserved quality Φ appears in the conditioning set, in contrast to more commonly used conditional orthogonality assumptions that condition on observed variables. Lastly, it is also worth clarifying that we only impose linearity in the price function, not on the anti-IV variable H.

2.2.2 Main Identification Result

The following lemma provides the foundation for our main theorem.

Lemma 2.5. *Consider a setting where home price is a linear function of amenity and unobserved quality (Assumption 2.1). If an anti-instrumental variable H satisfies conditional orthogonality (Assumption 2.4), then*

$$\tilde{H}(P, Z) = \tilde{H}(\Phi(P, Z)). \tag{2.8}$$

Proof.

$$\begin{aligned}
\tilde{H}(P, Z) &= \mathbb{E}^*[H|P, Z] && \text{by definition} && (2.9)\\
&= \mathbb{E}^*[H|P, Z, \Phi(P, Z)] && \text{by linear inversion } \Phi(P, Z) = P - \beta_0 - \beta_z Z && (2.10)\\
&= \mathbb{E}^*[H|Z, \Phi(P, Z)] && \text{by linearity of } P(Z, \Phi) && (2.11)\\
&= \mathbb{E}^*[H|\Phi(P, Z)] && \text{by conditional orthogonality} && (2.12)\\
&= \tilde{H}(\Phi(P, Z)) && \text{by definition} && (2.13)
\end{aligned}$$

□

In other words, for a given (p, z), Lemma 2.5 implies that $\tilde{H}(\Phi(p, z))$, the linear predicted value of H after projecting on *unobserved* $\Phi(p, z)$, can be obtained through the linear predicted anti-IV $\tilde{H}(p, z)$, which projects on *observed* price and amenity.

Theorem 2.6. *(Identification in the Linear Model) Consider a setting where home price is a linear function of amenity and unobserved quality (Assumption 2.1). If there exists a variable H that satisfies relevance (Assumption 2.3) and conditional orthogonality (Assumption 2.4),*

then, under mild regularity conditions such that the linear projection exists, the price of the amenity is identified as

$$\beta_z = -\frac{\delta_z}{\delta_p} \tag{2.14}$$

where δ_p and δ_z denote the linear projection coefficients of H on P and Z, respectively.

Proof. See Appendix A.[6] □

To gain some intuition for the identification result, recall that we are interested in identifying the marginal change in price P resulting from a marginal change in the amenity Z, while holding the unobserved quality Φ fixed. Relevance ensures that Φ maps one-to-one to $\tilde{H}(\Phi)$, so it is sufficient to hold $\tilde{H}(\Phi)$ fixed. The combination of the linear price and conditional orthogonality implies that $\tilde{H}(\Phi(P, Z)) = \tilde{H}(P, Z)$, as shown in Lemma 2.5. As a result, we just need to hold $\tilde{H}(P, Z)$ fixed to identify the implicit price of the amenity.

Figure 3 then provides the graphical intuition behind the ratio formula in Eq (2.14). Consider any point (p_0, z_0) with corresponding unobserved quality $\phi_0 = \Phi(p_0, z_0)$. The amenity price is represented by the tangent direction of the quality contour line, that is, the set of all the properties that share the same unobserved quality. Although the contour line with quality ϕ_0 is unobserved, we can substitute it with the contour line with the predicted anti-IV at $\tilde{H}(\phi_0)$, which is the same as $\tilde{H}(p_0, z_0)$. As such, the gradient vector of the predicted anti-IV $(\partial\tilde{H}/\partial Z, \partial\tilde{H}/\partial P)' = (\delta_z, \delta_p)'$ provides the direction of increasing quality, whereas the amenity price is represented by the vector perpendicular to the gradient, given by a slope of $-\delta_z/\delta_p$.

Thus, the predicted anti-IV $\tilde{H}(P, Z)$ captures the variation in H that is informative about unobserved quality, and once we condition on it, the remaining variation in Z provides us with the identifying variation for the implicit price.

Comparison with Conventional IV

It is worth discussing the connection and the difference between the anti-IV approach and the conventional linear IV approach. In particular, both approaches require a relevance assumption and an exclusion restriction to achieve identification, but the placement of these assumptions is reversed.

Table 1 succinctly summarizes the comparison: A conventional instrumental variable, denoted by IV, needs to be excluded from the structural error Φ, but be relevant for the endogenous variable of interest X^{EN}. By contrast, an anti-instrumental variable, denoted by AIV, is relevant for the structural error Φ, but conditionally excluded from the endogenous variable

[6] Note that the proof in Appendix A is readily generalizable to non-parametric price functions, laid out in Supplemental Appendix D. We also provide an alternative proof in B.1 that is more direct and relies explicitly on properties of the linear projection operator.

of interest X^{EN}. Consequently, the anti-IV method leverages an entirely different source of identifying variation compared with conventional IVs, and thus provides a new path for possible empirical strategies.

Table 1: A Comparison of IV vs. AIV (Linear Model)

	Instruments		**Anti-Instruments**	
Φ	*Exclusion*	$cov(IV, \Phi) = 0$	*Relevance*	$cov(AIV, \Phi) \neq 0$
X^{EN}	*Relevance*	$cov(IV, X^{EN}) \neq 0$	*Exclusion*	$cov(AIV^{\perp\Phi}, X^{EN}) = 0$

In addition, conditional orthogonality naturally nests the special case when $AIV = \Phi + \epsilon$, where ϵ is i.i.d. classical measurement error. In that case, one can also interpret the anti-IV approach as a strategy to deal with classical measurement error in the unobserved error, analogous to conventional IV being used as a way to deal with classical measurement error in the observed amenity.

2.2.3 Ratio-of-Coefficients Estimator

Theorem 2.6 also provides a constructive approach to estimate the amenity price.

Definition 2.7. (Ratio-of-Coefficients Estimator) Consider observations (H_i, Z_i, P_i). Let a first-stage OLS estimator be given by

$$H_i = \hat{\delta}_0 + \hat{\delta}_p P_i + \hat{\delta}_z Z_i + \mu_i. \tag{2.15}$$

The ratio-of-coefficients estimator is defined as

$$\hat{\beta}_z = -\frac{\hat{\delta}_z}{\hat{\delta}_p}. \tag{2.16}$$

Theorem 2.8. *(Consistency and Asymptotic Normality) Under the assumptions in Theorem 2.6, the ratio-of-coefficients estimator is consistent:*

$$\hat{\beta}_z = -\frac{\hat{\delta}_z}{\hat{\delta}_p} \xrightarrow{p} \beta_z. \tag{2.17}$$

Under additional mild regularity conditions, it is asymptotically normal:

$$\sqrt{n}(\hat{\beta}_z - \beta_z) \xrightarrow{d} \mathcal{N}(0, G'\Sigma G), \tag{2.18}$$

where Σ denotes the asymptotic covariance of the OLS estimator $\hat{\delta}$, and $G = (0, -1/\delta_p, -\delta_z/\delta_p^2)'$.

Proof. See Supplemental Appendix B.2. □

Weak Anti-IV

Just as conventional IV may encounter the problem of weak instruments (Andrews, Stock, and Sun, 2019), there could also be an analogous weak anti-instrument problem, because the

price coefficient δ_p from the linear projection $\mathbb{E}^*[H|P,Z]$ appears in the denominator of the implicit price $\beta_z = -\delta_z/\delta_p$. In particular, when δ_p is close to zero relative to the sampling variability of $\hat{\delta}_p$, the normal approximation to the distribution of the ratio becomes poor. To address this, we derive a corresponding Anderson-Rubin confidence interval that remains robust to weak identification (see Supplemental Appendix B.3).

Extension to Multiple Amenities Identification and estimation can be readily extended to accommodate a vector of amenities. Specifically, we allow them to be divided into a subset of focal amenities $Z_1, \ldots, Z_K$ and a subset of non-focal amenities used as controls D. If the corresponding conditional orthogonality assumption is satisfied for each focal amenity

$$\forall k = 1, \ldots, K : \; cov\left(H^{\perp \Phi, D}, Z_k\right) = 0, \tag{2.19}$$

then, their prices are identified by the ratio-of-coefficients as follows

$$\mathbb{E}^*[H|P, Z_1, ..., Z_K, D] = \delta_0 + \delta_p P + \delta_1 Z_1 + ... + \delta_K Z_K + \delta_D D \tag{2.20}$$

$$\forall k = 1, \ldots, K : \; \beta_k = -\frac{\delta_k}{\delta_p}. \tag{2.21}$$

A ratio-of-coefficients estimator for focal amenities is constructed analogously. The proof of the result follows directly from the proof of Theorem 2.6 in Supplemental Appendix B.1 by including all amenities Z_k and the controls D in the first-stage linear projection. Note that because the conditional orthogonality assumption in Eq (2.19) includes both D and Φ in the conditioning set, while the implicit price on Z_k is identified, the implicit price on D is not.

2.2.4 2SLS Estimator

As long as H is relevant (so that $\tilde{H}(\Phi)$ is a non-degenerate function of Φ), then Lemma 2.5 means that it is as good to control for $\tilde{H}(P,Z)$ as it is to control for $\tilde{H}(\Phi(P,Z))$. As such, we can use the predicted anti-IV $\tilde{H}(P,Z)$ as a "first-stage", which can then be used to identify the implicit price in a "second-stage."

Definition 2.9. (2SLS Estimator) Consider observations (H_i, Z_i, P_i). Let a first-stage OLS estimator be given by

$$H_i = \hat{\delta}_0 + \hat{\delta}_p P_i + \hat{\delta}_z Z_i + \mu_i. \tag{2.22}$$

The 2SLS estimator is obtained by the OLS regression of P on Z and $\hat{H}$ as the "second-stage"

$$P_i = \hat{\phi}_0 + \hat{\phi}_z Z_i + \hat{\phi}_{\hat{h}} \hat{H}_i + \omega_i. \tag{2.23}$$

Then, we let $\hat{\theta}_{2\text{SLS}} \equiv \hat{\phi}_z$.

It is straightforward to show that $\theta_{2\text{SLS}}$ is consistent (see Supplemental Appendix B.4). Conceptually, the 2SLS estimator can be interpreted as a control function estimator, namely,

we can project price P on amenity Z as long as we control for the linear predicted anti-IV $\tilde{H}(P,Z) = \mathbb{E}^*[H|P,Z]$. While the control function implementation that relies on conventional IVs use the *residual* from the first stage regression in the second stage (Wooldridge, 2007), the control function implementation of the anti-IV approach instead uses the *predicted value* from the first stage regression in the second stage, again highlighting that the anti-IV approach relies on a different source of identifying variation from conventional IVs.

2.2.5 GMM Estimator

We also formulate a GMM estimator, which allows us to handle both the just-identified case and the over-identified case when multiple anti-IVs are available. In particular, we re-formulate conditional orthogonality as the following moment conditions

$$g_i(\theta) = \begin{bmatrix} (P_i - \beta_z Z_i - \theta_h H_i - \theta_0) Z_i \\ (P_i - \beta_z Z_i - \theta_h H_i - \theta_0) P_i \\ (P_i - \beta_z Z_i - \theta_h H_i - \theta_0) \end{bmatrix} \tag{2.24}$$

where $\theta = (\beta_z, \theta_h, \theta_0)'$ include the key parameters of interest β_z. In Supplemental Appendix C, we provide relevant details for the GMM estimator, including consistency, asymptotic normality, and the construction of the corresponding weighting matrix for efficient estimation.

For easy implementation, we provide a Stata package *aivreg*, which includes all the estimators derived in this section, including the default *aivreg* for the ratio-of-coefficients estimator, *aivreg 2sls* for the "2SLS" estimator, and *aivreg gmm* for the GMM estimator.

2.3 Identification in the Non-Parametric Model

In this subsection, we develop the anti-IV approach for the general setting where the hedonic price function $P(Z,\Phi)$ is a non-parametric function of observed amenity Z and unobserved quality Φ. We do not impose any functional form restrictions, and in particular, we do not require separability or additivity in the unobserved quality.

2.3.1 Assumptions

Our non-parametric generalization does require one structural assumption placed on the hedonic pricing function, namely, the single index assumption.

Assumption 2.10. ***(Single Index)*** The hedonic pricing function $P(Z,\Phi)$ exists and is strictly monotone in a single index of unobserved quality Φ.

The single index assumption is an *economic* assumption about the underlying structure of how market prices are determined. We do not view this as a particularly restrictive assumption in the context of hedonic models, where Bajari and Benkard (2005) establish the existence of

a single-index hedonic pricing function only requires that the unobserved quality enter each household's utility with the same sign and nothing else.

Functionally, the single index assumption ensures that the pricing function is invertible in Φ. That is, there exists a unique inverse function $\Phi(P, Z)$ such that $P(Z, \Phi(P, Z)) = P$. Such invertibility serves a similarly important role in achieving non-parametric identification, analogous to the linear invertibility under the linear setting.

Next, consider an anti-instrument variable H. With a slight abuse of notation, we define *predicted anti-IV* $\tilde{H}(P, Z)$ as the conditional expectation of H conditioning on observed price P and amenity Z. Correspondingly, we $\tilde{H}(\Phi)$ conditions on unobserved quality Φ.

Definition 2.11. (Predicted Anti-IV)

$$\tilde{H}(P, Z) \equiv \mathbb{E}[H \mid P, Z] \tag{2.25}$$

$$\tilde{H}(\Phi) \equiv \mathbb{E}[H \mid \Phi] \tag{2.26}$$

Here, the conditional expectation operator $\mathbb{E}[H|P, Z]$ can be viewed as the non-parametric counterpart to the linear projection $\mathbb{E}^*[H|P, Z]$, which plays a central role in identification in the non-parametric case.

Next, we state the econometric assumptions required for non-parametric identification.

Assumption 2.12. ***(Relevance)*** In an open neighborhood around ϕ,

$$\tilde{H}(\phi) \text{ is strictly monotone in } \phi. \tag{2.27}$$

While stated as "monotonicity," we still consider it as a type of relevance condition, as it requires that the anti-IV H is, in expectation, informative of the unobserved quality, at least locally. In particular, it guarantees a one-to-one mapping between the predicted anti-IV $\mathbb{E}[H|\Phi]$ and unobserved quality Φ locally. Thus, the relevance assumption here can be thought of as the non-parametric analogue to the relevance assumption in Assumption 2.3, which guarantees a linear one-to-one mapping between the linear predicted anti-IV and unobserved quality.

Assumption 2.13. ***(Conditional Independence)*** An anti-instrumental variable H is conditionally independent of Z if

$$H \perp Z \mid \Phi. \tag{2.28}$$

Intuitively, conditional independence requires that the component of H unrelated to Φ be independent of Z. Note that, while we have relaxed the functional form restriction on price, we have strengthened the exclusion restriction from conditional orthogonality (Assumption 2.4) under the linear setting to conditional independence (Assumption 2.13) under the non-parametric setting.

2.3.2 Non-Parametric Identification

Theorem 2.14. ***(Non-Parametric Identification)*** *Consider a setting where the price function satisfies the single index assumption (Assumption 2.10). If there exists a variable H that satisfies relevance at ϕ (Assumption 2.12) and conditional independence (Assumption 2.13), then the price of the housing amenity is identified as*

$$\left.\frac{\partial P(Z,\Phi)}{\partial Z}\right|_{(z,\phi)} = -\frac{\partial \tilde{H}(P,Z)}{\partial Z} \Bigg/ \left.\frac{\partial \tilde{H}(P,Z)}{\partial P}\right|_{(p,z)}, \tag{2.29}$$

where $p = P(z,\phi)$, provided $P(z,\phi)$ is continuously differentiable at (z,ϕ), and $\tilde{H}(P,Z) = \mathbb{E}[H \mid P, Z]$ is well-defined at (p,z).

Proof. See Supplemental Appendix D. □

The intuition behind non-parametric identification is analogous to the linear setting. In particular, the relevance condition allows us to map one-to-one from unobserved quality Φ to the predicted anti-IV $\tilde{H}(\Phi)$. Then, the combination of conditional independence and the invertability of Φ, guaranteed by the single index assumption, yields $\tilde{H}(\Phi) = \mathbb{E}[H|\Phi] = \mathbb{E}[H|P,Z] = \tilde{H}(P,Z)$, which is the non-parametric counterpart to Lemma 2.5.[7] Therefore, it suffices to control for the predicted anti-IV $\tilde{H}(P,Z)$. Consequently, the implicit price can be represented as the ratio of partial derivatives in Eq (2.29), also analogous to the ratio of linear projection coefficients in Eq (2.14).

Thus, the non-parametric anti-IV approach can also be interpreted as a form of control function method in the spirit of Imbens and Newey (2009). Recall that it defines a variable V to be a control variable if it produces *conditional* independence between the variable of interest and the unobserved structural component, that is, $Z \perp \Phi \mid V$. In our setting, we can think of $V \equiv \tilde{H}(P,Z)$ as the control variable. To see why, note that

$$Z \perp \Phi \mid \Phi \qquad \text{by construction} \tag{2.30}$$

$$\iff Z \perp \Phi \mid \tilde{H}(\Phi) \qquad \textit{by local monotonicity} \tag{2.31}$$

$$\iff Z \perp \Phi \mid \tilde{H}(P(Z,\Phi),Z) \qquad \text{by } \textit{single index} \text{ and } \textit{cond. indep.} \text{ (Lemma D.1)} \tag{2.32}$$

It may seem a bit unusual that, once we control for the predicted anti-IV, no residual variation in Φ remains, as stated in Eq (2.30). However, this is exactly what the combination of the single index assumption and conditional independence, as demonstrated in the proof of Lemma D.1, enables us to leverage.

[7] See proof in Supplemental Appendix Lemma D.1.

2.3.3 Non-Parametric Estimation

Theorem 2.14 also provides a constructive approach to estimate the amenity price. One implementation involves producing a smooth non-parametric first-stage estimate of the conditional expectation function $\tilde{H}(P, Z) = \mathbb{E}[H|P, Z]$.[8] The second stage then computes the negative ratio of its partial derivatives to recover amenity prices. Alternatively, using the control function interpretation above, once we have an estimate of the first-stage predicted anti-IV, we can use it directly as a control in a second-stage estimate of the price function. Under appropriate regularity conditions, they generate consistent estimates.

2.3.4 Relation to Existing Methods

A methodological contribution of the anti-IV approach is that it highlights the critical role of conditional orthogonality (Assumption 2.4) or conditional independence (Assumption 2.13), which conditions on the *unobserved* component, in obtaining identification.

In studies of treatment effects, our anti-IV approach conceptually broadens the selection literature (Altonji, Elder, and Taber, 2005; Altonji and Mansfield, 2018; Oster, 2019; Masten and Poirier, 2024). Specifically, if an unobserved control in the selection framework satisfies the equal selection criterion, it also satisfies our conditional orthogonality assumption.[9] Unlike the selection literature which focuses on equal selection among the set of unobservable controls that contribute to the outcome, the anti-IV approach can leverage variables that are external to the structural equation itself as long as conditional independence is satisfied. In fact, the candidate anti-IV used in our empirical application, reported income, is not a housing feature that enters directly into the price function, yet it can still plausibly serve as an anti-IV.

In the estimation of the production function, our set-up with the single index assumption, which delivers the invertibility of unobserved quality, shares the same scalar invertibility property of the productivity residual. Here, the main difference is that the commonly used proxy-variable-based approach for production function estimation (Olley and Pakes, 1996; Levinsohn and Petrin, 2003) requires a perfect mapping between the proxy variable and unobserved productivity (conditioning on observables), which also makes it subject to functional form dependence or requiring timing restrictions (Ackerberg, Caves, and Frazer, 2015). On the other hand, the anti-IV approach does not seek to take a stance on the structural relationship required to generate a perfect proxy for unobserved productivity, but we only rely on a variable that is on average informative. In turn, we rely on conditional independence as the econometric property required in delivering identification.

The conditional independence condition has also appeared in the causal inference literature.

[8] There is a large body of work devoted to non-parametric estimation in general, including kernel, local linear regression, and series/sieve approaches (Fan and Gijbels, 1992; Newey, 1997; Li and Racine, 2006; Chen, 2007; Belloni, Chernozhukov, and Hansen, 2014).

[9] See derivations in Supplemental Appendix B.5.

However, because those settings do not impose single-index as the structural relationship, they must combine conditional independence with other restrictions to attain identification. For example, the proximal causal inference framework in epidemiology (Miao, Geng, and Tchetgen Tchetgen, 2018; Tchetgen, Ying, Cui, Shi, and Miao, 2020; Deaner, 2023) requires two set of auxiliary variables, negative control outcomes and negative control treatments, to be independent conditional on the unobserved confounder. Our framework does not need the second set of negative controls related to the treatment. Separately, De Luna, Fowler, and Johansson (2017) also formalize such conditional independence requirement but impose an additional proxy property, requiring the potential outcome be independent of the confounder conditional on the proxy, which we do not impose. Thus, our contribution is to show that when single-index is an appropriate structural assumption, which may be especially relevant for the hedonic model given Bajari and Benkard (2005), a relevant anti-IV that satisfies conditional independence can also deliver non-parametric identification.

3 Background and Data

Our main empirical application of the anti-instrument is to estimate the implicit price of property-level flood risk. In this section, we describe the various sources used to construct a combined dataset of housing characteristics, flood risk measures, and household demographics, all at the individual property level and covering housing transactions in the US from 2018 to 2023. Table 2 presents summary statistics for the key variables.

Property Transactions and Attributes (ATTOM)

We obtain a nationwide property-level dataset from local tax assessor records collected by ATTOM, a third-party company that provides detailed information on property attributes as well as transaction price, transaction date, and mortgage information.[10] Variables from the assessor files provide the exact geocoded location of each property as well as information on a broad set of housing characteristics, including the number of bedrooms/bathrooms, square footage, lot size, and build year. We restrict our sample of transactions to those that involve an arm's length transaction with a transacted price between $20,000 and $5M to exclude outliers in the tails of the distribution.

Property-Level Flood Risk (First Street and FEMA)

We consider two measures of flood risk: First Street Flood Factor and FEMA floodplains.

First Street Flood Factor The First Street Flood Factor is a forward-looking, predictive measure of future flood risk at the parcel-level, developed by First Street Technology (formerly

[10] ATTOM excludes non-public disclosure states, namely, ID, KS, MS, MT, NM, TX, UT, WY, in terms of real estate prices.

First Street Foundation). A key feature of this model is that it incorporates simulated natural disasters and flood events over the next 30 years based on meteorological projections under various climate scenarios.[11]

Our data come from First Street Flood Model V2, which includes a composite measure of future flood risk assigned to the centroid of each parcel's building footprint. It produces a score ranging from 1 (lowest risk) to 10 (highest risk), capturing both the likelihood and depth of flooding. Table 3 breaks down the transaction data by the property's Flood Factor. While over 85% of all sold properties have a minimal risk score of 1, a substantial fraction still faces material flood risk. For example, about 1.9% of homes sold have "extreme" flood risk with a Flood Factor of 9 or 10, which typically has over a 98% likelihood of flooding over the next 30 years.

We merge the ATTOM property database with the First Street dataset by matching each parcel in the transaction data to the nearest parcel in the First Street data based on latitude and longitude coordinates of the parcel centroids. This spatial merge assigns a unique flood risk measure from First Street to each property in our sample.

FEMA Floodplains As part of the data distribution, First Street also provides information on parcel-level location within flood zones designated by the Federal Emergency Management Agency (FEMA). An area is considered to be in the "100-Yr floodplain" if it is expected to be "inundated by the flood event having a 1-percent chance of being equaled or exceeded in any given year," also known as the Special Flood Hazard Area (SFHA).[12] Within the 100-Yr floodplain, National Flood Insurance Program (NFIP) enforces floodplain management regulation and requires mandatory flood insurance for homes purchased with federally backed or regulated mortgages.[13]

Correspondingly, a property is in the "500-Yr floodplain" if it is located in an area with a 0.2% annual chance of flooding, indicating lower flood risk than the 100-Yr floodplain. Properties that fall within the 500-Yr floodplain but outside the 100-Yr floodplain are not subject to mandatory flood insurance requirements.

Flood risks measured using the First Street Flood Factors and the FEMA floodplains are generally correlated. Table 3 shows that properties with higher First Street Flood Factors are generally more likely to be located in the 100-Yr floodplain. For example, over half of the properties with a Flood Factor of 10 are in the 100-Yr floodplain. However, not all properties

[11]For more details on the First Street Foundation Flood Model, see https://firststreet.org/methodology/flood.

[12]Flood Zones, https://www.fema.gov/glossary/flood-zones, accessed July 5, 2024.

[13]While the mandatory flood insurance requirement applies only to federally backed or regulated mortgages, in practice, most lenders extend this requirement to all residential mortgages in the 100-year floodplain (Ge, Lam, and Lewis, 2022).

in the 100-Yr floodplain have high Flood Factors; some are assigned relatively low Flood Factor scores.[14]

For the purpose of our analysis, we treat both of them as plausible measures of flood risk for which we seek to estimate implicit prices. Due to the mandatory insurance requirement, it is usually salient to the home buyer whether a property is located in the 100-Yr floodplain, whereas First Street Flood Factor only became widely accessible on Realtor.com in August 2020 and was later adopted by other major listing platforms.[15] For our purpose, it is not essential that home buyers know the exact First Street Flood Factor; rather, we use Flood Factor as a measure of the underlying flood risk which is partially known through a combination of local knowledge and flood disclosures.

Household Income and Demographics (HMDA)

We combine property transactions with the Modified Loan Application Register (MLAR), the public version of LAR submitted by financial institutions under the Home Mortgage Disclosure Act (HMDA). It allows us to capture income and demographic characteristics of the mortgage applicant(s), including race, ethnicity, gender, and age at the time of their mortgage application.

In particular, the HMDA income field captures the "gross annual income" of the mortgage applicant, reported to the lending institution in making its credit decision. Unlike other sources of self-reported income, this figure is verified by the lender and is required by GSEs for a completed mortgage.[16] While income alone does not fully determine a borrower's ability to repay their loan, the verification process ensures that the reported amount is at least a reliable measure of borrower income at the time of mortgage application.

Since HMDA does not include transaction-level identifiers, following Bayer, Ferreira, and Ross (2016) and Bernstein, Billings, Gustafson, and Lewis (2022), our matching process is based on the combination of transaction year, lender name, sales price, loan amount, and census tract.[17] At the end, about 55% of nationwide transactions match uniquely to HMDA records. Among unmatched transactions, approximately 60% are associated with cash purchases that did not involve a mortgage. However, a priori, it seems unlikely that cash buyers are strongly selected

[14]The differences likely stem from a combination of methodology and spatial precision. FEMA floodplain designations are area-level classifications based primarily on land elevation contour lines and statistical analysis of historical flood events; they are not inherently forward-looking (National Research Council, 2009). In contrast, First Street Flood Factors are property-level assessments that account for finer geographic variation and explicitly incorporate risks associated with climate-driven changes.

[15]Realtor.com Press Release (August 26, 2020) Now Available - Flood Risk Data for Homes. Redfin.com Press Release (Feb 17, 2021) Redfin Publishes Flood Risk Data from Flood Factor for Over 94 Million Homes, accessed June 5, 2025.

[16]Lenders typically verify income using a combination of pay stubs, tax forms, and other documentation. See detailed procedures for various income sources (e.g., primary and secondary employment, commission, rental income) from Fannie Mae and Freddie Mac, accessed June 5, 2025.

[17]Since 2018, HMDA reports loan amounts and sales prices as midpoints of $10,000 bins. We round loan amounts and sales prices in the transaction data to the same midpoints when matching to HMDA.

on flood risk preferences, and even if they are, it still seems unlikely that the marginal price of flood risk is determined solely by cash buyers, with no mortgage borrowers in the support.

Table 2 summarizes our main dataset after merging housing transactions with property-level measures of flood risk and household income. The final sample contains 8,198,017 property records. The median home price (in 2020 dollars) is $325,000, and the median annual household income is $90,000.

Additional Data Sources

We supplement our property-level attributes with spatial data from the U.S. Geological Survey (USGS) to capture elevation and distance to the coast. Parcel-level elevation is measured using centroid-based values from USGS digital elevation models at 10-meter resolution. We also compute Euclidean distance to the coastline based on USGS waterline maps.

In addition, we augment household demographic information using voter registration data. Specifically, we use a nationwide voter file from L2 covering approximately 190 million individuals. For each household, we assign a political affiliation score based on the median registered individual: Republican (+1), Democrat (–1), or Independent (0).[18] Because our voter file reflects registration status as of November 2020, we only observe political affiliation for the subset of households whose home purchase occurred prior to that date.

4 The Anti-IV Assumptions in Pricing Flood Risk

In this section, we discuss whether the key assumptions underlying the anti-IV approach can be applied to estimating the implicit price of flood risk.

4.1 The Single Index Assumption

Underlying the anti-IV approach is a *structural* assumption that the hedonic pricing function is determined by observed amenities and is strictly monotonic in a single index of unobserved quality. Is this assumption reasonable? What restrictions does the single index assumption impose on underlying housing demand and supply? Theorem 1 of Bajari and Benkard (2005) establishes the conditions that guarantee both the existence of the equilibrium price function $P(Z, \Phi)$ and the single-index property in a differentiated product market. It shows that the existence of a single-index hedonic pricing function imposes no restrictions on the supply side

[18] For a household where the median individual is between a Republican (+1) and a Democrat (-1), we assign it 0.

and only minimal conditions on the demand side for any hedonic model[19] of consumer utility. The only substantive requirement is that the unobserved characteristic Φ enters all household utility functions with the same sign. It is in this sense that the unobserved characteristic is a *vertical* attribute, meaning all households agree on the *sign* with which it enters their utility. Intuitively, if the unobserved component were horizontal, invertibility may fail, as multiple values of the unobserved component could rationalize the same observed price and amenity combination. Also, without loss of generality, we can focus on the case when price is increasing in the unobserved component, thus interpreting Φ as quality.

While it may seem restrictive to concatenate all unobservables into a single quality index, we believe it to be a reasonable approximation for the housing context. Regressions using observable controls already explain a large share of the price variation (e.g., about 90% in our application). The remaining unobserved component captures a combination of hard-to-measure attributes of a home, such as build quality, maintenance, curb appeal, interior finishes, the condition of structural components, etc. To the extent that our concern is unobserved quality being correlated with the focal amenity and leading to biased estimation, it is precisely this single-dimensional, vertically differentiated component that we seek to capture.[20] Moreover, in the existing literature, both Bajari, Fruehwirth, Kim, and Timmins (2012) and Diamond and Diamond (2024) have relied on the single index assumption to address unobserved quality in the housing markets, albeit with a different focus on leveraging panel data.

That said, there will always be scope for missing amenities, rendering the hedonic model misspecified. However, to the extent that such omitted features can be absorbed into the definition of the unobserved quality, we still obtain unbiased estimate provided that the anti-IV satisfies the requisite conditional independence assumption with respect to that adjusted quality. If this condition still fails, the resulting bias can be characterized by a formula based on the joint correlation between the omitted amenity and the unobserved quality.[21] In our application, because we measure flood risk at the individual property level and include extremely granular

[19]By "hedonic," we refer to any model of individual utility that does not include an idiosyncratic error term $\epsilon_{i,j}$ specific to both the individual i and the choice j. While such $\epsilon_{i,j}$ terms are often included in discrete choice models, they are often introduced for computational convenience to account for further heterogeneous preferences over unobserved product characteristics and to rationalize positive market shares for observably dominated products. Pure characteristics models without $\epsilon_{i,j}$ are generally more difficult to estimate (Berry and Pakes, 2007; Agarwal, 2015). However, because our focus is on estimating implicit prices rather than matching quantities, requiring the underlying utility to be hedonic remains an appropriate representation of the choice environment.

[20]It is also worth clarifying that, while Epple, Quintero, and Sieg (2020) and Landvoigt, Piazzesi, and Schneider (2015) also consider a single vertical dimension of differentiation of the housing market, our goal is to control for such vertical differences as a confounder, which is entirely distinct from their approach that collapses all housing and neighborhood attributes into a single quality index as a focal subject of inquiry, which, at the time, methods for uncovering unobserved heterogeneity in preferences were only developed for markets with a single product attribute (Ekeland, Heckman, and Nesheim, 2004; Heckman, Matzkin, and Nesheim, 2010).

[21]We derive this bias expression in Supplemental Appendix E.

geographic fixed effects at the census block level, the concern about omitted neighborhood amenities is less relevant for estimating the price of flood risk.

Lastly, imposing a single index of unobserved quality is not more restrictive than existing approaches in most empirical studies. In essence, when researchers use quasi-experiments and consider linear regressions that include an error term, they implicitly assume a single dimension of the unobservable component.

4.2 Micro-Founding Reported Income as Anti-IV

Next, we discuss plausible candidates for anti-instrumental variables. To price flood risk under a linear hedonic setting, we illustrate that, within a class of housing choice models, a measure of household budget such as reported income at the time of mortgage application, can be micro-founded as both relevant for unobserved quality and plausibly conditionally orthogonal to amenity choices.

A Stylized Model of Housing Choice

To begin, consider an indirect utility specification for household i and housing unit j that is log-linear in characteristics:

$$u_{i,j} = \alpha_i P_j + \gamma_{i,1} \ln(Z_{j,1}) + \ldots + \gamma_{i,K} \ln(Z_{j,K}) + \ln(\Phi_j) \tag{4.1}$$

We allow households to have heterogeneous price sensitivities α_i, heterogeneous preferences $\gamma_{i,k}$ for amenities $Z_{j,k}$, and a single dimension of unobserved quality Φ_j. In addition, we allow price sensitivity α_i to flexibly depend on household budget η_i, for example,

$$\alpha_i = \alpha \eta_i^{\lambda - 1}, \tag{4.2}$$

where λ denotes the Box-Cox parameter. When $\lambda = 0$, we have $\alpha_i = \alpha / \eta_i$, which approximates the canonical Cobb-Douglas utility specification, as used in Berry, Levinsohn, and Pakes (1995). In the more general case, Miravete, Seim, and Thurk (2024) shows that allowing flexibility in λ is key to more accurately approximating the true shape of demand.

While the price sensitivity term α_i is typically modeled as a function of household income (often denoted as y_i) in the discrete choice literature, we emphasize that, in our setting, we consider the driver of price sensitivity in home purchase decisions as a function of the household's underlying budget, η_i, which can be viewed as a composite measure reflecting household income, assets, liquidity, expected future income, etc; although the true household budget may be unobserved to the econometrician, we may have access to a relevant measure of it, such as the reported income at the time of mortgage application.

As such, the utility specification in Eq (4.1) is parsimonious yet flexible. It allows for heterogeneity in preferences over both prices and amenities and permits unobserved quality to be

correlated with them. The key restriction is that, conditional on price and the vector of observed amenities, there is no additional heterogeneity in preferences for the unobserved quality Φ_j. In other words, differences in willingness-to-pay for unobserved housing quality are driven solely by variation in price sensitivity α_i. This is a widely used assumption in the empirical literature: it is imposed on Φ_j in virtually all empirical models of differentiated products following Berry, Levinsohn, and Pakes (1995) (where it is commonly denoted ξ_j), including many structural models of housing markets (Bayer, Ferreira, and McMillan, 2007; Galiani, Murphy, and Pantano, 2015; Calder-Wang, 2021; Cook, Li, and Binder, 2024). Importantly, the plausibility of this assumption crucially depends on whether the included amenities are rich enough to capture sufficient preference heterogeneity. In our housing application that includes a broad set of housing and neighborhood characteristics, we believe this is a plausible assumption.

Relevance and Conditional Independence

Now, we consider a linear hedonic model of the housing market. To simplify exposition, we assume the single-amenity case with $K = 1$. Utility maximization of Eq (4.1) then yields the following first-order condition:

$$\frac{\partial u_i}{\partial Z} = \alpha_i \frac{\partial P}{\partial Z} + \gamma_i \frac{1}{Z} = 0 \implies Z_i^* = -\frac{1}{\alpha_i} \frac{\gamma_i}{\partial P/\partial Z} = -\frac{1}{\alpha_i} \frac{\gamma_i}{\beta_z}, \tag{4.3}$$

$$\frac{\partial u_i}{\partial \Phi} = \alpha_i \frac{\partial P}{\partial \Phi} + \frac{1}{\Phi} = 0 \implies \Phi_i^* = -\frac{1}{\alpha_i} \frac{1}{\partial P/\partial \Phi} = -\frac{1}{\alpha_i}. \tag{4.4}$$

where the second set of equalities relies on the linearity assumption $P = \beta_0 + \beta_z Z + \Phi$.

The optimal choice of amenity Z_i^* is determined by both a household's price sensitivity α_i and the strength of their preference for the amenity γ_i, scaled by its price β_z. At the same time, because we have assumed no additional preference heterogeneity over unobserved quality, the optimal choice of quality Φ_i^* depends only on α_i, scaled by the price of quality, which has been normalized to 1. Since we have modeled price sensitivity as a function of household budget, $\alpha_i = \alpha \eta_i^{\lambda - 1}$, the resulting choice of unobserved quality maps directly to the size of the household budget:

$$\Phi_i^* = -\eta_i^{1-\lambda}/\alpha. \tag{4.5}$$

Now consider a potential anti-IV variable H_i, such as the reported income of the primary mortgage applicant. Because reported income is likely increasing on average with the true household budget, it is also likely increasing on average with the chosen level of unobserved housing quality, thereby satisfying the *relevance* condition. Said differently, knowing something about household characteristics, particularly their income at the time of mortgage application, can be particularly informative about the unobserved quality of the home they live in.

More importantly, the conditional orthogonality condition amounts to

$$cov\left(H_i, Z_i^* \,\middle|\, \Phi_i^*\right) = 0 \iff cov\left(H_i, \frac{1}{\alpha_i}\frac{\gamma_i}{\beta_z} \,\middle|\, \frac{1}{\alpha_i}\right) = 0 \iff cov\left(H_i, \frac{\eta_i^{1-\lambda}}{\alpha}\frac{\gamma_i}{\beta_z} \,\middle|\, \frac{\eta_i^{1-\lambda}}{\alpha}\right) = 0 \tag{4.6}$$

$$\iff cov\left(H_i, \eta_i^{1-\lambda}\gamma_i \,\middle|\, \eta_i^{1-\lambda}\right) = 0 \iff cov\left(H_i, \gamma_i \,\middle|\, \eta_i\right) = 0. \tag{4.7}$$

In other words, the main conditional orthogonality boils down to requiring Eq (4.7), that is, conditional on the true household budget η_i, reported income H_i is uncorrelated with the household's amenity preference γ_i. Conceptually, the housing choice model allows us to swap unobserved housing quality Φ_i^* with unobserved household budget η_i in the conditioning set.

Intuitively, reported applicant income may mismeasure one's true budget for purchasing a house due to various factors. For example, gross income in a given year does not account for deductions or debt obligations, it omits variations in household net assets or liquidity, and it may reflect year-to-year idiosyncratic fluctuations. The conditional orthogonality assumption, in this context, means that these mismeasurements are not correlated with household preferences for flood risk.

Additionally, to the extent that a household's preference for flood risk may be correlated with its budget η_i unconditionally (for example, wealthier households may be more likely to care about flood risk), this is entirely permissible in our framework. The key requirement is *conditional* orthogonality, not total orthogonality, between reported income and preferences for the amenity.

Overall, the key insight delivered by the model is that if households with larger budgets choose greater amount of unobserved housing quality, then a relevant measure of household budget, such as reported income at the time of mortgage application, can serve as an anti-IV for identifying the implicit amenity price, provided that the mis-measurement of the budget is uncorrelated with flood preference.

Limitation and Discussion

The purpose of this discussion is to illustrate that there exists a class of models in which a measure of household budget can be micro-founded as both relevant for unobserved quality and plausibly conditionally uncorrelated with amenity choices. Within this class of models, there is a direct one-to-one relationship that maps the true household budget η_i to the chosen unobserved quality Φ_i^*, which then allows us to swap the conditioning set from unobserved quality to household budget.

That said, in practice, researchers may readily consider other models of the underlying utility and prices where such a direct relationship between household budget and housing quality does

not hold.[22] Since the one-to-one relationship is a sufficient condition for swapping out the conditioning set, its absence does not necessarily imply a violation of the conditional orthogonality assumption.

In either case, one strategy to directly assess whether conditional orthogonality is appropriate is to check whether the estimated coefficients remain robust after controlling for observable correlates of amenity preferences or observable correlates of residual variation in the anti-IV conditional on unobserved quality. For example, if one's political affiliation (or age, race, ethnicity) is believed to be correlated with one's amenity preference, then one would want to assess how the estimated coefficient is affected by the inclusion of such demographic correlates of preferences in the estimation. Relatedly, if we have observable correlates that are informative of the residual variation in the anti-IV after conditioning on unobserved quality (e.g., observable correlates of how reported income may mismeasure unobserved housing quality in non-random ways), then we can also consider adding those controls. As with any exclusion restrictions, one can never prove that the condition holds. However, if the estimated coefficient remains robust after controlling for these observable correlates, this would increase our confidence in its validity.

5 Empirical Results

In this section, we apply the anti-IV approach to estimate the implicit price of flood risk.

5.1 Main Estimates

Table 5 shows our main estimates of log sales price on Flood Factor indicators (2–10), using Flood Factor 1 as the omitted category. Column (1) starts with the standard OLS regression controlling for county fixed effects, year fixed effects, and a vector of housing characteristics, including indicators of bathroom counts, building-age in decade bins, log square feet, and log lot size. In this OLS specification, all coefficients for Flood Factors above 4 appear with a positive sign and are increasing with the level of flood risk. The coefficient on flood factor 10 is a positive 11.6%, implying that, on average, homes with the highest projected flood risk appear 11.6% more expensive, which is wrong-signed. In Columns (2)-(4), we introduce increasingly granular geographic fixed effects including census tract, census block group, and census block, and observe that the coefficient in front of Flood Factor 10 falls to 4.6%, 3.6%, and 2.4%, respectively, but remains significantly positive. Figure 2 panel (a) visualizes these wrong-signed price estimates.

[22]For instance, without the first-order Taylor approximation of the Box-Cox specification, it would require preferences for the numeraire good to adjust so that the expenditure share of unobserved quality remains the same among households with the same budget. An example of it can be found in Eq (E.26) in Supplemental Appendix E.4. Separately, if prices are linear in logs, it would require the underlying utility to be in log prices.

In particular, starting column (4), we include census block fixed effects, allowing us to hold constant a wide array of neighborhood-level characteristics. A census block is the smallest U.S. Census unit, typically a city block bounded by streets,[23] capturing local amenities such as school quality, labor market access, transit, retail, and environmental factors like air quality or heat risk.

In Column (5), we include income reported from HMDA as a control variable in the conventional OLS specification. Here, reported income enters strongly with a significantly positive coefficient, but the coefficient on the highest Flood Factor 10 is still a positive 2.1%, showing that simply controlling for reported income still does *not* resolve the wrong-sign problem.

In Column (6), we use the anti-IV approach. We perform a first-stage regression of log reported income on log home price, Flood Factor, and other controls, then use the ratio-of-coefficients estimator from Definition 2.7 to compute the implicit prices. We find insignificant estimates for low to moderate risk categories (from 2 to 8). At Flood Factor 9, the estimated coefficient becomes statistically significant at -0.66%. At Flood Factor 10, the estimated coefficient is statistically significant at -3.48%, suggesting that homes with a Flood Factor 10 sell for 3.48% *lower* than homes with a Flood Factor 1, all else equal. In addition, we do not have a weak anti-IV problem, with the partial $\mathcal{F}$ associated with the inclusion of price in the first stage regression extremely high (over 400,000) and a partial Rsq of about 6.6%, suggesting the denominator δ_p is significantly different from 0. As such, the weak-inference robust ARCI coincides entirely with the asymptotic estimates in this case.

Figure 2 panel (b) visualizes our price estimates using the anti-IV approach. Strikingly, the anti-IV estimates decline sharply at the highest risk levels, precisely the opposite of the positive, rising pattern produced by OLS even under the most granular controls. Moreover, the anti-IV estimator does not "know" the numeric order of the Flood Factor labels, but it is still able to recover a declining pattern in price as risk increases.

Assessing Anti-IV Validity

The key exclusion restriction of the anti-IV approach is that, conditional on unobserved quality, reported income is uncorrelated with the chosen level of flood risk. Equivalently, through the lens of the housing choice model in Section 4.2, the exclusion restriction becomes that, conditional on the true household budget, which determines the chosen unobserved quality, how reported income mismeasures true budget must be uncorrelated with a household's flood-risk preference.

What could be potential reasons that conditional orthogonality may be violated? There could be various reasons why reported income may mismeasure the true household budget that is non-random, and if that non-randomness is also correlated with one's flood preference, then it would

[23] What are census blocks? *US Census Bureau*. Retrieved July 4, 2024.

constitute a violation. For example, Black or Hispanic households likely have lower wealth than Whites for a given income (Derenoncourt, Kim, Kuhn, and Schularick, 2024), which implies that minority households likely have higher reported income for a given underlying budget. If minority status is also correlated with preferences for flood risk, this would breach conditional orthogonality. To assess its potential impact, Table 6 column (1) includes an indicator for Black applicants and another indicator for Hispanic applicants. We find that the coefficients for Flood Factors 9 and 10 remain essentially unchanged, suggesting no substantive bias due to correlations with race.

Another driver of the mismeasurement of household budget by reported income could be age. Older households often have more assets relative to current income, so reported income may understate their true budget. If age also affects one's flood-risk preference, this too would violate the exclusion restriction. In Column (2), we include five HMDA age-cohort dummies (< 35, $35-44$, $45-54$, $55-64$, $65+$), and we see the anti-IV estimates for high-risk factors shift negligibly (the estimate for Flood Factor 10 moves only from -3.48% to -3.32%), suggesting correlation with age has minimal impact on the empirical results.

Additionally, if researchers have observable correlates of amenity preferences, then adding them as a control can also be a direct way to assess conditional independence (or conditional orthogonality). In our setting, political affiliations may be correlated with one's preference for flood risk due to different perceptions of climate risk (Bernstein, Billings, Gustafson, and Lewis, 2022). If the extent to which reported income mismeasures one's true budget is also systematically different between Republicans and Democrats (although this seems rather unlikely), then this would also constitute a violation. In column (3), we restrict to the subsample with which we have voter registration data, and in column (4), we include the household's median party affiliation as a control. The resulting anti-IV estimates again remain unchanged between these two specifications, lending support for conditional orthogonality in our setting.

Lastly, as discussed in Section 4.1, in a setting with multiple amenities, one might also be concerned about the impact of missing amenities, which is a violation of the single index assumption.[24] While we think the scope of missing amenities is likely limited with the inclusion of block-fixed effect in our setting, we may still be concerned about the correlation between the residual variations in elevation or distance-to-coast within a block (which are amenities due to better views (Bin, Crawford, Kruse, and Landry, 2008)) and property-level flood risk, even conditioning on unobserved quality. Table 6 columns (5) and (6) show that flexibly control for bins of elevation and distance-to-coast has no material impact on the estimated price of flood

[24] A violation of the single index assumption can also be viewed as a violation of the conditional independence in this case as the quality term becomes mis-specified.

risk.[25]

Non-parametric Estimates

Because the anti-IV approach non-parametrically identifies implicit prices, we can produce local estimates of the amenity price without imposing any functional form assumptions on the hedonic price function $P(Z, \Phi)$. While there are various approaches at producing the first-stage non-parametric estimate of the predicted anti-IV function $\tilde{H}(P, Z) = \mathbb{E}[H|P, Z]$, as a proof of concept, we illustrate a partition-based implementation.[26]

As shown in Figure 4, we divide all transactions into 11 $\times$ 10 cells, each representing homes within a certain home price quantile (1-11) and Flood Factor category (1-10). We approximate the predicted anti-IV function $\tilde{H}(P, Z)$ by taking the average of the anti-IV H in each cell, after partialing out block-group fixed effects. The top panel displays a heat-map of $\tilde{H}$, where lighter colors indicate higher predicted incomes. The similarly colored cells trace out a downward-sloping contour, suggesting a negative price. In the bottom panel, we recover a local implicit price at each cell by finite-difference approximations $-\frac{\partial \tilde{H}/\partial Z}{\partial \tilde{H}/\partial P}$, computed using predicted income differences one bin apart. Consistent with our linear anti-IV results, the highest-risk categories exhibit the largest negative prices.

5.2 Estimates of FEMA Floodplain

While our main analysis focuses on using property-level First Street Flood Factor to measure risk, we also apply the anti-IV approach to estimate the implicit price of being in a FEMA floodplain. The advantage of FEMA's 100-yr floodplain designation is that homeowners are likely aware of their status via mandatory insurance requirement; the disadvantage is that the floodplain designation is coarse in that a property is either in the floodplain or not, making it a noisy measure of the underlying flood risk.

Table 4 summarizes our findings. Column (1) presents a conventional OLS with census block fixed effects and house-level controls, producing a wrong-signed 0.73% premium for being in the 100-Yr floodplain. Adding reported income as a control in column (2) reduces this to 0.49%, but still with a positive sign. In column (3), using reported income as an anti-IV yields a correctly signed -2.9% implicit price, implying a *discount* for properties in the 100-Yr floodplain. We do not find statistically significant results for being in the 500-Yr floodplain.

A negative implicit price of the 100-Yr floodplain is sensible given the mandatory insurance requirement. As a back-of-the-envelope calculation, FEMA reports a median NFIP premium

[25] To allow for non-linearity, we divide distance-to-coast into bins of $[0, 0.05)$, $[0.05, 0.1)$, $[0.1, 0.25)$, $[0.25, 0.5)$, $[0.5, 2)$ and more than 2 miles. We divide elevation into bins of $[0, 1)$, $[1, 2)$, ..., $[9, 10)$, $[10, 50)$, $[50, 100)$, $[100, 500)$ and higher than 500 meters.

[26] In general, methods such as local polynomial regression are more suitable for estimating a derivative. For the purpose of illustration, we use a partition-based implementation here because our measure of flood risk, Flood Factor, is categorical in nature.

of $1,290 per year;[27] discounted at 8% (Poterba and Solomon, 2021) over an average length of home-ownership of nine years (Mateyka and Marlay, 2010), the total cost of premium amounts to a present value of $8,058. For a median home priced at $325,000, our estimate of 2.9% equals $9,360, which is quite close. Existing research leveraging changes in the NFIP policies (Georgic and Klaiber, 2022; Hennighausen, Liao, Nolte, and Pollack, 2023; Ge, Lam, and Lewis, 2022) also finds that mandatory flood insurance costs are capitalized into housing prices with broadly comparable magnitudes. That said, we still caution against doing a direct comparison of the insurance costs with the estimated amenity price due to inherent conceptual differences, as the implicit price is driven by the marginal owner's willingness to pay rather than the costs incurred by the average.

5.3 Beyond Flood Risk

While the focus of this paper is to illustrate how to apply the anti-IV method in estimating the price of flood risk, in theory, the proposed method could also help estimate the price of other residential amenities, especially other environmental amenities that are also known to be prone to the wrong-signed problem. However, as discussed before, the anti-IV approach, while well-suited to address issues related to the unobserved vertical quality, could still lead to biased estimates if there exist missing amenities that are otherwise required by the single index assumption.

In other words, for amenities with limited variation within a geographic area (e.g., air pollution, heat, fire, etc.), we become inherently constrained in our ability to include granular geographic fixed effects, and we have to lean more heavily on a carefully constructed anti-IV where conditional orthogonality could plausibly hold with respect to the corresponding unobserved quality.[28] Lastly, we leave the question of how to extend the framework to the case of multiple dimensions of unobservables to future research.

6 Conclusion

In this paper, we develop a novel anti-instrumental variable approach that allows us to estimate the implicit price of amenities in the housing market in the midst of unobserved quality. Our estimator can be applied to commonly used observational datasets, and proves that the typical confounder of unobserved vertical quality that is to blame for classic "wrong-signed" results should no longer be viewed as endemic to observational approaches. We show

[27] Cost of Flood Insurance for Single Family Homes under NFIP's Pricing Approach, Exhibit 2, accessed July 4, 2024.

[28] That said, as a proof of concept, Bell, Calder-Wang, and Zhong (2024) develop a "Geographic PageRank" to be an alternative measure of place quality based on migration flows, which can then be used as an anti-IV to obtain plausible estimates of the price of air pollution that are broadly in line with the literature (Chay and Greenstone, 2005). Hence, applying the method more broadly to other residential amenities would require a careful evaluation of the exclusion restriction with respect to the focal amenity in question.

that the implicit prices can be non-parametrically identified under a notably different set of assumptions, central to which is that the researcher has available to her an "anti-instrument" that is relevant for unobserved quality but conditionally orthogonal / independent of the focal amenity.

To investigate whether such a viable anti-instrument might exist in the context of pricing flood risk, we leverage reported income at the time of mortgage application to help us account for unobserved housing quality. Through the lens of a housing choice model where a household's chosen amount of quality is driven by its budget set, the identifying assumption can be viewed as requiring the mismeasurement of true budget by reported income to be uncorrelated with one's flood preference. The anti-IV approach produces statistically significant estimates whereby the marginal American home-buyer pays 3.5% *less* to own a property with the highest flood risk at a Flood Factor of 10. The magnitude of our estimate also monotonically increases with the level of projected flood risk. In a literature where finding credible estimates of flood risk has been challenging, our estimates will prove relevant for understanding the economic impact of flood risk, or more broadly, climate change, on residential real estate, a major segment of the economy.

The anti-instrumental variable approach is not without limitations. While the presence of a hedonic pricing function that satisfies the single index assumption could be a sensible approximation to the housing equilibrium, it may not apply in other settings, especially when there are multiple unobserved amenities or when a significant portion of prices are driven by idiosyncratic variation. Moreover, it also depends on the availability of a candidate anti-instrument that can plausibly satisfy conditional independence.

Overall, we believe that the anti-IV approach provides an important tool in a researcher's arsenal to address endogeneity issues that arise naturally when the outcome of interest is jointly determined by observed characteristics and an unobserved vertical component. If anything, given the prevalence of traditional instrumental variable approaches, researchers have always been looking for variables that can instrument for the endogenous, *observed* characteristic. Rather, our research shines a spotlight on a previously overlooked area, where we suggest searching for candidate variables that may instrument for the endogenous, *unobserved* component, opening up new sources of identification for these problems.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83(6), 2411–2451.

Agarwal, N. (2015): "An Empirical Model of the Medical Match," *American Economic Review*, 105(7), 1939–1978.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113(1), 151–184.

Altonji, J. G., and R. K. Mansfield (2018): "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects," *American Economic Review*, 108(10), 2902–2946.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11(Volume 11, 2019), 727–753.

Bajari, P., and C. L. Benkard (2005): "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*.

Bajari, P., J. C. Fruehwirth, K. I. Kim, and C. Timmins (2012): "A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution," *American Economic Review*, 102(5), 1898–1926.

Baldauf, M., L. Garlappi, and C. Yannelis (2020): "Does Climate Change Affect Real Estate Prices? Only If You Believe In It," *The Review of Financial Studies*, 33(3), 1256–1295.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115(4), 588–638.

Bayer, P., F. Ferreira, and S. L. Ross (2016): "The Vulnerability of Minority Homeowners in the Housing Boom and Bust," *American Economic Journal: Economic Policy*, 8(1), 1–27.

Bell, A. (2020): "Job Amenities and Earnings Inequality," *Working Paper*.

Bell, A., S. Calder-Wang, and S. Zhong (2024): "Measuring Housing Quality Using Revealed Preference: A Geographic PageRank Approach," *Working Paper*.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, 28(2), 29–50.

Beltrán, A., D. Maddison, and R. Elliott (2019): "The Impact of Flooding on Property Prices: A Repeat-Sales Approach," *Journal of Environmental Economics and Management*, 95, 62–86.

Bernstein, A., S. B. Billings, M. T. Gustafson, and R. Lewis (2022): "Partisan Residential Sorting on Climate Change Risk," *Journal of Financial Economics*, 146(3), 989–1015.

Bernstein, A., M. T. Gustafson, and R. Lewis (2019): "Disaster on the Horizon: The Price Effect of Sea Level Rise," *Journal of Financial Economics*, 134(2), 253–272.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S., and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48(4), 1193–1225.

Bin, O., T. W. Crawford, J. B. Kruse, and C. E. Landry (2008): "Viewscapes and Flood Hazard: Coastal Housing Market Response to Amenities and Risk," *Land Economics*, 84(3), 434–448.

Bin, O., and C. E. Landry (2013): "Changes in Implicit Flood Risk Premiums: Empirical Evidence from the Housing Market," *Journal of Environmental Economics and Management*, 65(3), 361–376.

Birnbaum, M. H. (1979a): "Is There Sex Bias in Salaries of Psychologists?," *American Psychologist*, 34(8), 719–720.

——— (1979b): "Procedures for the Detection and Correction of Salary Inequities," *Salary equity*, pp. 121–144.

Bishop, K. C., N. V. Kuminoff, H. S. Banzhaf, K. J. Boyle, K. Von Gravenitz, J. C. Pope, V. K. Smith, and C. D. Timmins (2020): "Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality," *Review of Environmental Economics and Policy*, 14(2), 260–281.

Bishop, K. C., and C. Timmins (2019): "Estimating the Marginal Willingness to Pay Function without Instrumental Variables," *Journal of Urban Economics*, 109, 66–83.

Brown, J. N., and H. S. Rosen (1982): "On the Estimation of Structural Hedonic Price Models," *Econometrica*, 50(3), 765–768.

Burbano, V., O. Folke, S. Meier, and J. Rickne (2022): *The Gender Gap in Meaningful Work: Explanations and Implications*. Centre for Economic Policy Research.

Calder-Wang, S. (2021): "The Distributional Impact of the Sharing Economy on the Housing Market," *Working Paper*.

Calvin, K., D. Dasgupta, G. Krinner, A. Mukherji, P. W. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, W. W. Cheung, S. Connors, F. Denton, A. Diongue-Niang, D. Dodman, M. Garschagen, O. Geden, B. Hayward, C. Jones, F. Jotzo, T. Krug, R. Lasco, Y.-Y. Lee, V. Masson-Delmotte, M. Meinshausen, K. Mintenbeck, A. Mokssit, F. E. Otto, M. Pathak, A. Pirani, E. Poloczanska, H.-O. Pörtner, A. Revi, D. C. Roberts, J. Roy, A. C. Ruane, J. Skea, P. R. Shukla, R. Slade, A. Slangen, Y. Sokona, A. A. Sörensson, M. Tignor, D. Van Vuuren, Y.-M. Wei, H. Winkler, P. Zhai, Z. Zommers, J.-C. Hourcade, F. X. Johnson, S. Pachauri, N. P. Simpson, C. Singh, A. Thomas, E. Totin, P. Arias, M. Bustamante, I. Elgizouli, G. Flato, M. Howden, C. Méndez-Vallejo, J. J. Pereira, R. Pichs-Madruga, S. K. Rose, Y. Saheb, R. Sánchez Rodríguez, D. Ürge-Vorsatz, C. Xiao, N. Yassaa, A. Alegría, K. Armour, B. Bednar-Friedl, K. Blok, G. Cissé, F. Dentener, S. Eriksen, E. Fischer, G. Garner, C. Guivarch, M. Haasnoot, G. Hansen, M. Hauser, E. Hawkins, T. Hermans, R. Kopp, N. Leprince-Ringuet, J. Lewis, D. Ley, C. Ludden, L. Niamir, Z. Nicholls, S. Some, S. Szopa, B. Trewin, K.-I. Van Der Wijst, G. Winter, M. Witting, A. Birt, M. Ha, J. Romero, J. Kim, E. F. Haites, Y. Jung, R. Stavins, A. Birt, M. Ha, D. J. A. Orendain, L. Ignon, S. Park, Y. Park, A. Reisinger, D. Cammaramo, A. Fischlin, J. S. Fuglestvedt, G. Hansen, C. Ludden, V. Masson-Delmotte, J. R. Matthews, K. Mintenbeck, A. Pirani, E. Poloczanska, N. Leprince-Ringuet, and C. Péan (2023): "IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland.," Discussion paper, Intergovernmental Panel on Climate Change (IPCC).

Chay, K. Y., and M. Greenstone (2005): "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 113(2), 376–424.

Chen, X. (2007): "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, vol. 6, pp. 5549–5632. Elsevier.

Chernozhukov, V., A. Galichon, M. Henry, and B. Pass (2021): "Identification of Hedonic Equilibrium and Nonseparable Simultaneous Equations," *Journal of Political Economy*, 129(3), 842–870.

Conway, D. A., and H. V. Roberts (1983): "Reverse Regression, Fairness, and Employment Discrimination," *Journal of Business & Economic Statistics*, 1(1), 75–85.

Cook, C., P. Li, and A. Binder (2024): "Where to Build Affordable Housing? Evaluating the Tradeoffs of Location," *Working Paper*.

De Luna, X., P. Fowler, and P. Johansson (2017): "Proxy Variables and Nonparametric Identification of Causal Effects," *Economics Letters*, 150, 152–154.

De Schouwer, T., and I. Kesternich (2024): "Work Meaning and the Flexibility Puzzle," *Working Paper*.

Deaner, B. (2023): "Proxy Controls and Panel Data," *Working Paper*.

Derenoncourt, E., C. H. Kim, M. Kuhn, and M. Schularick (2024): "Wealth of Two Nations: The U.S. Racial Wealth Gap, 1860–2020," *The Quarterly Journal of Economics*, 139(2), 693–750.

Diamond, R., and W. Diamond (2024): "Racial Differences in the Total Rate of Return on Owner-Occupied Housing," *Working Paper*.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(S1), S60–S109.

Epple, D. (1987): "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, 95(1), 59–80.

Epple, D., L. Quintero, and H. Sieg (2020): "A New Approach to Estimating Equilibrium Models for Metropolitan Housing Markets," *Journal of Political Economy*, 128(3), 948–983.

Fan, J., and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20(4), 2008–2036.

Ferreira, S., S. M. de Morentin, and A. Erro-Garcés (2025): "Measuring Job Risks When Hedonic Wage Models Do Not Do the Job," *Journal of Environmental Economics and Management*, 130, 103120.

Folke, O., and J. Rickne (2022): "Sexual Harassment and Gender Inequality in the Labor Market," *The Quarterly Journal of Economics*, 137(4), 2163–2212.

Galiani, S., A. Murphy, and J. Pantano (2015): "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment," *American Economic Review*, 105(11), 3385–3415.

Ge, S., A. Lam, and R. Lewis (2022): "The Effect of Insurance Premiums on the Housing Market and Climate Risk Pricing," *Working Paper*.

Georgic, W., and H. A. Klaiber (2022): "Stocks, Flows, and Flood Insurance: A Nationwide Analysis of the Capitalized Impact of Annual Premium Discounts on Housing Values," *Journal of Environmental Economics and Management*, 111, 102567.

Gibson, M., and J. T. Mullins (2020): "Climate Risk and Beliefs in New York Floodplains," *Journal of the Association of Environmental and Resource Economists*, 7(6), 1069–1111.

Giglio, S., B. Kelly, and J. Stroebel (2021): "Climate Finance," *Annual Review of Financial Economics*, 13(Volume 13, 2021), 15–36.

Goldberger, A. S. (1984): "Reverse Regression and Salary Discrimination," *The Journal of Human Resources*, 19(3), 293.

Hansen, B. (2022): *Econometrics*. Princeton University Press.

Hashimoto, M., and L. Kochin (1980): "A Bias in the Statistical Estimation of the Effects of Discrimination," *Economic Inquiry*, 18(3), 478–486.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010): "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5), 1569–1591.

Hennighausen, H., Y. Liao, C. Nolte, and A. Pollack (2023): "Flood Insurance Reforms, Housing Market Dynamics, and Adaptation to Climate Risks," *Journal of Housing Economics*, 62, 101953.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

Kapsalis, C. (1982): "A New Measure of Wage Discrimination," *Economics Letters*, 9(3), 287–293.

Kuminoff, N. V., C. F. Parmeter, and J. C. Pope (2010): "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities?," *Journal of Environmental Economics and Management*, 60(3), 145–160.

Kuminoff, N. V., and J. C. Pope (2014): "Do "Capitalization Effects" for Public Goods Reveal the Public's Willingness to Pay?," *International Economic Review*, 55(4), 1227–1250.

Landvoigt, T., M. Piazzesi, and M. Schneider (2015): "The Housing Market(s) of San Diego," *American Economic Review*, 105(4), 1371–1407.

Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70(2), 317–341.

Li, Q., and J. Racine (2006): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Maideu-Morera, G. (2024): "Amenity-Based Technical Change," *Working Paper*.

Masten, M. A., and A. Poirier (2024): "The Effect of Omitted Variables on the Sign of Regression Coefficients," .

Mateyka, P., and M. Marlay (2010): "The Duration and Tenure of Residence, 1996 to 2009," *Census Working Papers*.

Miao, W., Z. Geng, and E. J. Tchetgen Tchetgen (2018): "Identifying Causal Effects with Proxy Variables of an Unmeasured Confounder," *Biometrika*, 105(4), 987–993.

Miravete, E. J., K. Seim, and J. Thurk (2024): "Elasticity and Curvature of Discrete Choice Demand Models," *Working Paper*.

Murfin, J., and M. Spiegel (2020): "Is the Risk of Sea Level Rise Capitalized in Residential Real Estate?," *The Review of Financial Studies*, 33(3), 1217–1255.

National Research Council (2009): "Mapping the Zone: Improving Flood Map Accuracy," Discussion paper, National Research Council, Washington, D.C.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147–168.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Ortega, F., and S. Taşpinar (2018): "Rising Sea Levels and Sinking Property Values: Hurricane Sandy and New York's Housing Market," *Journal of Urban Economics*, 106, 81–100.

Oster, E. (2019): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37(2), 187–204.

Poterba, J. M., and A. Solomon (2021): "Discount Rates, Mortality Projections, and Money's Worth Calculations for US Individual Annuities," *NBER Working Paper*.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34–55.

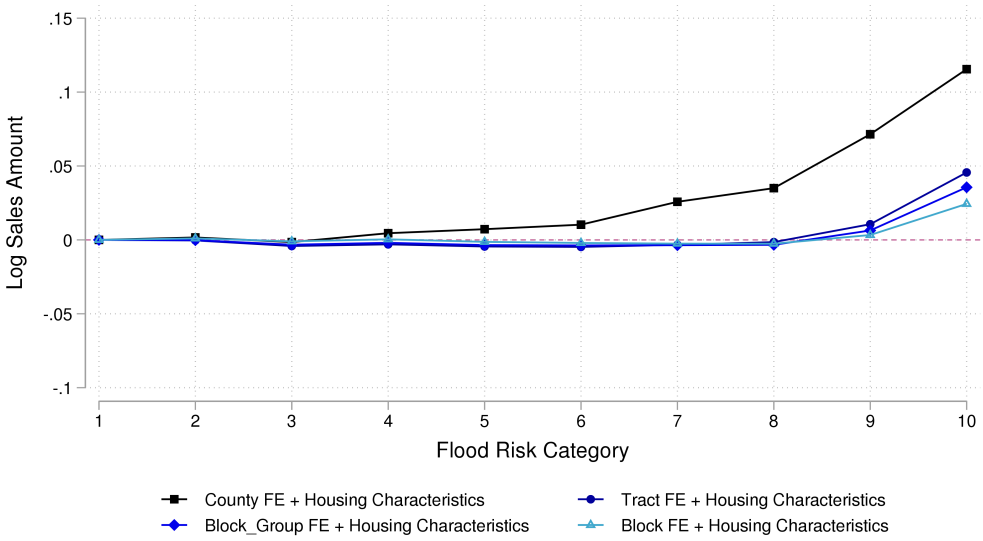
Tchetgen, E. J. T., A. Ying, Y. Cui, X. Shi, and W. Miao (2020): "An Introduction to Proximal Causal Learning," *arXiv Preprint*.

Wooldridge, J. (2007): "What's New in Econometrics? Lecture 6 Control Functions and Related Methods," *NBER Methods Lecture*.

7 Figures

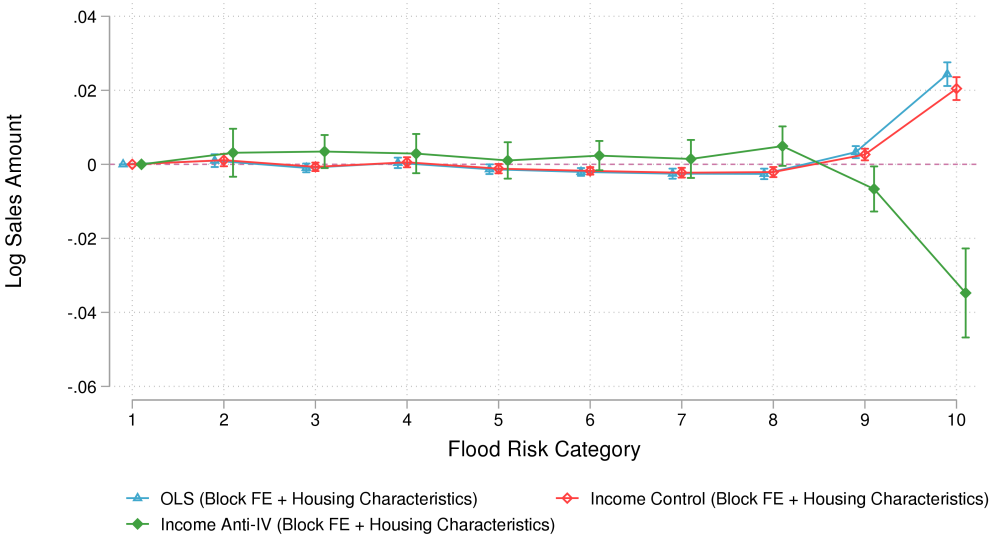
Figure 2: Pricing Flood Risk: OLS and Anti-IV

(a) Illustration of the Wrong-Sign Problem with OLS



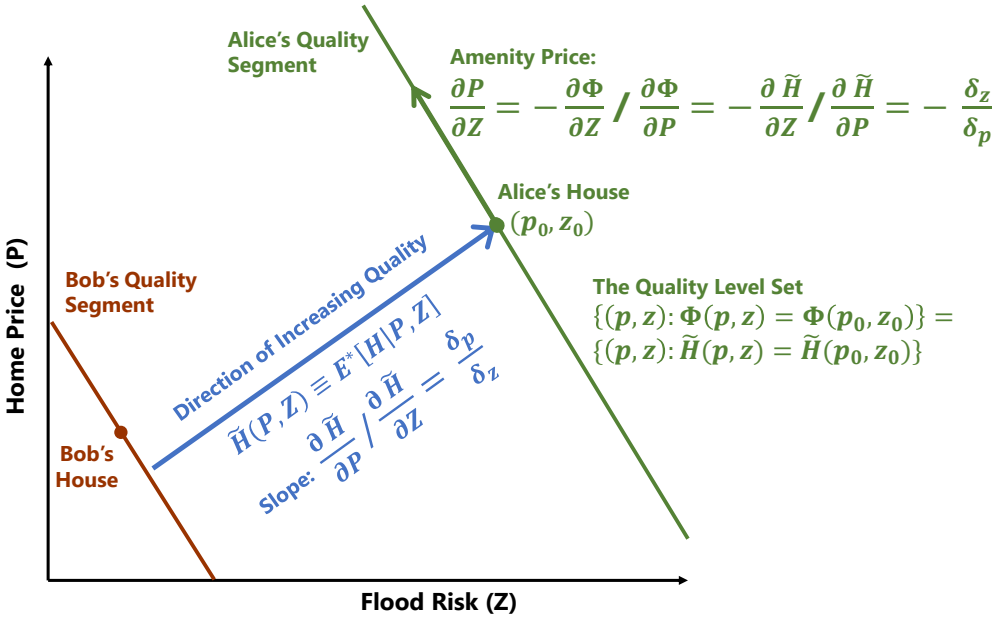
In panel (a), we plot the coefficients when regressing log home prices on various Flood Risk categories (2-10), with 1 being the omitted category, using home sales data from 2018 to 2023. We include controls for various property characteristics and geographic fixed effects, including county, census tract, census block-group, and census block fixed effects, respectively.

(b) Pricing Flood Risk Using Anti-IV



In panel (b), we begin with the OLS regression with property characteristics controls and census block fixed effects. We subsequently add controls for reported income. Lastly, we use reported income as an anti-IV, which shows a significant and negative estimate of the price of flood risk. Moreover, the magnitude also increases with the underlying risk among high-risk properties.

Figure 3: Graphical Illustration of Identification Using the Anti-IV Approach



Notes: The figure above illustrates the intuition behind the identification formula. By definition, the price of the amenity at any given observed is determined by how prices change in response to a change in amenity, holding its unobservable quality fixed. Thus, the amenity price at (p_0, z_0) can be represented by the tangent direction of the level set

$$\{(p,z) : \Phi(p,z) = \Phi(p_0,z_0)\} = \{(p,z) : \tilde{H}(\Phi(p,z)) = \tilde{H}(\Phi(p_0,z_0))\} \tag{7.1}$$

$$= \{(p,z) : \tilde{H}(p,z) = \tilde{H}(p_0,z_0)\} \tag{7.2}$$

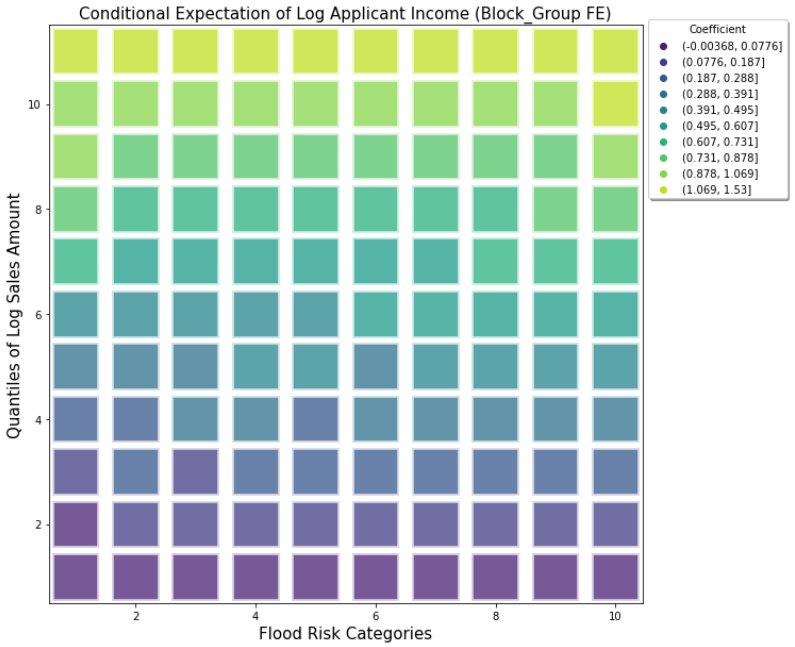
where the first equality is a result of the relevance of $\tilde{H}(\Phi)$, and the second equality is a result of Lemma 2.5, namely, $\tilde{H}(p,z) = \tilde{H}(\Phi(p,z))$. Thus, the amenity price can be estimated as the tangent vector of the level set of the predicted anti-IV function, namely,

$$\frac{\partial P}{\partial Z} = -\frac{\partial \Phi}{\partial Z} \Big/ \frac{\partial \Phi}{\partial P} = -\frac{\partial \tilde{H}}{\partial Z} \Big/ \frac{\partial \tilde{H}}{\partial P} = -\frac{\delta_z}{\delta_p}. \tag{7.3}$$

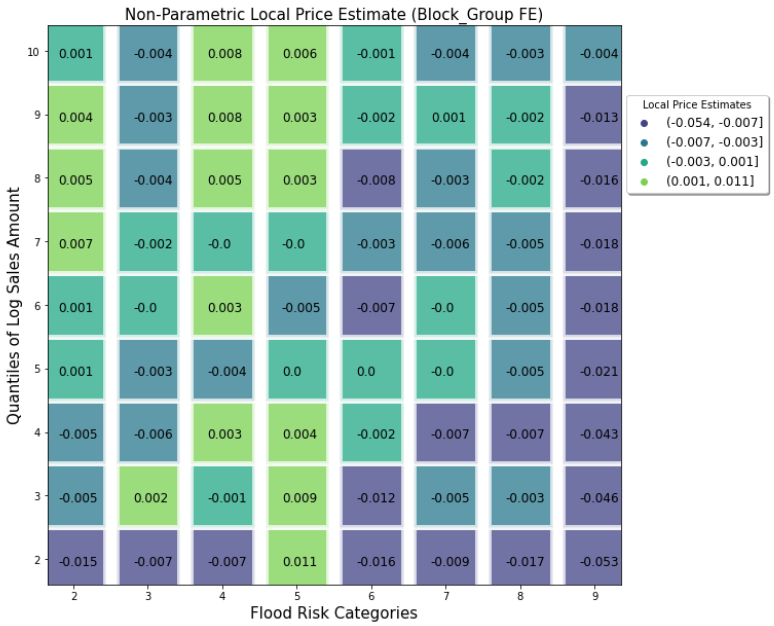
Moreover, the tangent vector is perpendicular to the gradient $\nabla \tilde{H} = (\partial \tilde{H}/\partial Z, \partial \tilde{H}/\partial P)' = (\delta_z, \delta_p)'$, which represents the direction of increasing quality.

Figure 4: Non-parametric Estimation of Amenity Prices

(a) Predicted Reported Income



(b) Local Price Estimates



Notes: This figure categorizes the homes into 110 (11x10) cells. Each cell represents homes that lie within certain house price quantile and flood factor category. The top panel plots the predicted anti-IV function, computed as the average reported income in that cell, controlling for block-group fixed effects. The bottom panel plots the estimated price of a one-category increase in the Flood Factor.

8 Tables

Table 2: Summary Statistics

	N	Mean	Stdev	p10	p25	p50	p75	p90
Property Transactions and Characteristics (ATTOM):								
Sales Price (in $)	8,198,017	405,503.15	332,802.58	155,000.0	215,000.0	325,000.0	475,000.0	715,000.0
Year of Sales	8,198,017	2,020.16	1.50	2,018.0	2,019.0	2,020.0	2,021.0	2,022.0
Year Built	8,107,556	1,979.45	31.58	1,938.0	1,959.0	1,986.0	2,004.0	2,017.0
Total Square Ft	8,198,017	1,951.19	933.49	1,085.0	1,344.0	1,762.0	2,356.0	3,054.0
Lot Size Square Ft	8,198,017	46,183.11	2,997,432.66	4,025.0	6,445.0	9,744.0	18,066.0	44,431.0
Number of Bathrooms	8,198,017	2.25	1.16	1.0	2.0	2.0	3.0	4.0
Household Demographics (HMDA):								
Applicant Income (in $1,000s)	8,198,017	117.62	380.31	43.0	60.0	90.0	136.0	207.0
Age <34	8,198,017	0.39	0.49	0.0	0.0	0.0	1.0	1.0
Age 35–44	8,198,017	0.28	0.45	0.0	0.0	0.0	1.0	1.0
Age 45–54	8,198,017	0.17	0.37	0.0	0.0	0.0	0.0	1.0
Age 55–64	8,198,017	0.10	0.30	0.0	0.0	0.0	0.0	1.0
Age >65	8,198,017	0.05	0.22	0.0	0.0	0.0	0.0	0.0
Is Black	8,198,017	0.08	0.27	0.0	0.0	0.0	0.0	0.0
Is Hispanic	8,198,017	0.02	0.14	0.0	0.0	0.0	0.0	0.0
Additional Information:								
Elevation (meters)	8,196,315	243.60	346.64	8.5	44.1	173.0	280.0	426.0
Distance to Coast (miles)	8,198,017	124.99	168.86	5.1	14.1	50.9	182.3	318.3
Median Household Republican	4,634,732	0.24	0.43	0.0	0.0	0.0	0.0	1.0
Median Household Democrat	4,634,732	0.30	0.46	0.0	0.0	0.0	1.0	1.0

Notes: The combined sample is based on arm's length transactions of U.S. single family homes from 2018 to 2023 with sale prices between $20,000 and $5M. Sales price and reported applicant income are harmonized to inflation adjusted values to 2020 dollars.

Table 3: Summary Statistics: Flood Risk Measures

First Street Flood Factor	First Street Risk Category	Count	Fraction	% In 500-Yr Flood Zone (Not 100-Yr)	% In 100-Yr Flood Zone
1	Minimal	7,024,310	85.7%	4.06%	0.96%
2	Minor	76,652	0.9%	20.83%	7.92%
3	Moderate	169,423	2.1%	15.81%	7.42%
4	Moderate	124,637	1.5%	14.60%	8.71%
5	Major	144,575	1.8%	14.15%	10.70%
6	Major	248,028	3.0%	11.99%	14.27%
7	Severe	132,802	1.6%	12.90%	19.49%
8	Severe	126,958	1.6%	11.13%	23.76%
9	Extreme	114,423	1.4%	7.63%	39.90%
10	Extreme	36,209	0.4%	7.96%	53.17%
Total		8,198,017	100 %	5.36%	3.28%

Table 4: Estimation of the Price of Being in FEMA Floodplain

	Log Sales Amount		
	(1) OLS	(2) Control	(3) Anti-IV
100-Yr Flood Zone	0.00732***	0.00494***	-0.0288***
	(0.000747)	(0.000722)	(0.00281)
500-Yr Flood Zone (Not 100-Yr)	0.00145	0.00153	0.00272
	(0.000887)	(0.000857)	(0.00333)
Log Reported Income		0.116***	
		(0.000174)	
Geographic FEs	Block	Block	Block
Housing Characteristics	✓	✓	✓
Year FEs	✓	✓	✓
N	7539789	7539789	7539789
Adjusted R^2	0.909	0.915	
Partial R^2			0.0660
Partial $\mathcal{F}$			441839.2

Notes: The sample is based on arm's length transactions of U.S. single family homes from 2018 to 2023. Census block boundaries are defined using the 2010 census maps. The omitted category is Flood Factor 1. Housing characteristics controls include indicators of bathroom counts, building-age in decade bins (including a bin for missing building age), log square feet, and log lot size. All regressions are based on the same sample, with the total number of observations adjusted for singletons due to the inclusion of fixed effects. Standard errors in parentheses. Statistical significance at the 0.1%, 1%, and 5% levels are denoted by ***, **, and *, respectively.

Table 5: Estimation of the Price of Flood Risk

	Log Sales Amount					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) Control	(6) Anti-IV
Flood Factor=2	0.00166	-0.000370	-0.000190	0.000999	0.00114	0.00313
	(0.00108)	(0.000827)	(0.000817)	(0.000879)	(0.000850)	(0.00331)
Flood Factor=3	-0.00150*	-0.00413***	-0.00325***	-0.000966	-0.000674	0.00346
	(0.000732)	(0.000562)	(0.000556)	(0.000606)	(0.000586)	(0.00228)
Flood Factor=4	0.00454***	-0.00301***	-0.00197**	0.000394	0.000560	0.00291
	(0.000849)	(0.000658)	(0.000654)	(0.000718)	(0.000694)	(0.00270)
Flood Factor=5	0.00725***	-0.00446***	-0.00349***	-0.00132*	-0.00116	0.00105
	(0.000794)	(0.000617)	(0.000612)	(0.000668)	(0.000645)	(0.00251)
Flood Factor=6	0.0103***	-0.00473***	-0.00375***	-0.00205***	-0.00176***	0.00236
	(0.000614)	(0.000481)	(0.000479)	(0.000537)	(0.000519)	(0.00202)
Flood Factor=7	0.0258***	-0.00332***	-0.00360***	-0.00252***	-0.00226***	0.00146
	(0.000825)	(0.000641)	(0.000637)	(0.000696)	(0.000673)	(0.00262)
Flood Factor=8	0.0350***	-0.00150*	-0.00332***	-0.00259***	-0.00209**	0.00492
	(0.000850)	(0.000664)	(0.000658)	(0.000723)	(0.000699)	(0.00272)
Flood Factor=9	0.0715***	0.0106***	0.00639***	0.00330***	0.00265***	-0.00666*
	(0.000897)	(0.000728)	(0.000728)	(0.000827)	(0.000799)	(0.00311)
Flood Factor=10	0.116***	0.0456***	0.0356***	0.0244***	0.0205***	-0.0348***
	(0.00157)	(0.00130)	(0.00133)	(0.00163)	(0.00158)	(0.00614)
Log Reported Income					0.116***	
					(0.00174)	
Geographic FEs	County	Tract	Block Group	Block	Block	Block
Housing Characteristics	✓	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓	✓
N	8182529	8181231	8176940	7539789	7539789	7539789
Adjusted R^2	0.784	0.883	0.891	0.909	0.915	
Partial R^2						0.0660
Partial F						441819.7

Notes: The sample is based on arm's length transactions of U.S. single family homes from 2018 to 2023. Boundaries for census tract, block-group, and block are defined using the 2010 census maps. The omitted category is Flood Factor 1. Housing characteristics controls include indicators of bathroom counts, building-age in decade bins (including a bin for missing building age), log square feet, and log lot size. All regressions in the table are based on the same sample, with the total number of observations adjusted for singletons due to the inclusion of fixed effects. Standard errors in parentheses. Statistical significance at the 0.1%, 1%, and 5% levels are denoted by ***, **, and *, respectively.

Table 6: Robustness: Estimation of the Price of Flood Risk

	Log Sales Amount					
	Additional Demographic Characteristics		Political Affiliation (Subsample)		Additional Housing Features	
	(1)	(2)	(3)	(4)	(5)	(6)
Flood Factor = 2	0.00315	0.00381	0.00465	0.00479	0.00319	0.00348
	(0.00331)	(0.00331)	(0.00513)	(0.00514)	(0.00331)	(0.00331)
Flood Factor = 3	0.00343	0.00341	0.00662	0.00668	0.00349	0.00381
	(0.00228)	(0.00228)	(0.00358)	(0.00358)	(0.00228)	(0.00228)
Flood Factor = 4	0.00287	0.00269	-0.00258	-0.00257	0.00293	0.00342
	(0.00270)	(0.00270)	(0.00423)	(0.00424)	(0.00270)	(0.00271)
Flood Factor = 5	0.00101	0.00136	-0.00283	-0.00292	0.00107	0.00170
	(0.00251)	(0.00251)	(0.00393)	(0.00394)	(0.00251)	(0.00252)
Flood Factor = 6	0.00235	0.00297	0.00395	0.00382	0.00234	0.00311
	(0.00202)	(0.00202)	(0.00317)	(0.00318)	(0.00202)	(0.00203)
Flood Factor = 7	0.00144	0.00110	0.00628	0.00623	0.00148	0.00252
	(0.00262)	(0.00262)	(0.00408)	(0.00409)	(0.00262)	(0.00263)
Flood Factor = 8	0.00490*	0.00543*	0.00779	0.00797	0.00493	0.00624*
	(0.00272)	(0.00272)	(0.00425)	(0.00426)	(0.00272)	(0.00273)
Flood Factor = 9	-0.00669*	-0.00618*	-0.00515	-0.00524	-0.00643*	-0.00494
	(0.00311)	(0.00311)	(0.00487)	(0.00488)	(0.00311)	(0.00314)
Flood Factor = 10	-0.0348***	-0.0332***	-0.0440***	-0.0440***	-0.0342***	-0.0302***
	(0.00614)	(0.00613)	(0.00980)	(0.00982)	(0.00614)	(0.00622)
Geographic FEs	Block	Block	Block	Block	Block	Block
Housing Characteristics	✓	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓	✓
Additional Controls	Minority Indicator	Age Categories	None	Voter Affiliation	Miles to Coast	Elevation
N	7,539,789	7,539,789	3,971,092	3,971,092	7,539,789	7,538,301

Notes: The sample is based on arm's length transactions of U.S. single family homes from 2018 to 2023. Boundaries for census tract, block-group, and block are defined using the 2010 census maps. The omitted category is Flood Factor 1. Housing characteristics controls include indicators of bathroom counts, building-age in decade bins (including a bin for missing building age), log square feet, and log lot size. Minority indicators include indicators of whether the primary mortgage applicant is Black and whether the primary mortgage applicant is Hispanic. The dummies for age categories include < 35, $35-44$, $45-54$, $55-64$, $65+$. Controls for voter affiliation include indicators of whether the median household member is registered Republican, Democrat, or Independent. Controls for distance-to-coast are indicators of bins of $[0, 0.05)$, $[0.05, 0.1)$, $[0.1, 0.25)$, $[0.25, 0.5)$, $[0.5, 2)$ and more than 2 miles. Controls for elevation are indicators of bins of $[0, 1)$, $[1, 2)$, ..., $[9, 10)$, $[10, 50)$, $[50, 100)$, $[100, 500)$ and higher than 500 meters. Standard errors in parentheses. Statistical significance at the 0.1%, 1%, and 5% levels are denoted by ***, **, and *, respectively.

A Appendix: Proof of Identification

Proof of Identification: The Linear Case (Theorem 2.6)

Proof. First, the regularity conditions include finite moments ($\mathbb{E}[H^2] < \infty$, $\mathbb{E}\|X\| < \infty$) and sufficient variation ($Q_{XX} = \mathbb{E}[XX']$ is positive definite), where $X = (P, Z, 1)'$. This guarantees the linear projection $\mathbb{E}^*[H|P, Z]$ uniquely exists (Theorem 2.9, Hansen (2022)).

For any given (Z, Φ), Lemma 2.5 implies

$$\tilde{H}(P(Z, \Phi), Z) = \tilde{H}(\Phi) \tag{A.1}$$

Take the total derivative of the LHS of Equation (A.1)

$$\frac{d}{dZ}\Big(\tilde{H}(P(Z, \Phi), Z)\Big) = \frac{\partial \tilde{H}(P, Z)}{\partial P}\frac{\partial P(Z, \Phi)}{\partial Z} + \frac{\partial \tilde{H}(P, Z)}{\partial Z} = \delta_p \beta_z + \delta_z \tag{A.2}$$

Correspondingly, take the total derivative of the RHS of Equation (A.1)

$$\frac{d}{dZ}\Big(\tilde{H}(\Phi)\Big) = 0. \tag{A.3}$$

Taken together, we have $\delta_p \beta_z + \delta_z = 0$. Lastly, consider the denominator

$$\delta_p = \frac{\partial \tilde{H}(P, Z)}{\partial P} = \frac{\partial \tilde{H}(\Phi(P, Z))}{\partial \Phi} \times \frac{\partial \Phi(P, Z)}{\partial P} \neq 0, \tag{A.4}$$

because relevance ensures $\partial \tilde{H}(\Phi)/\partial \Phi \neq 0$ and linearity ensures $\Phi(P, Z) = P - \beta_0 - \beta_z Z$, and thus $\partial \Phi(P, Z)/\partial P \neq 0$. Therefore, the implicit price is identified as the ratio of the linear projection coefficients $\beta_z = -\delta_z/\delta_p$.

□

B Supplemental Appendix: Additional Proofs

B.1 Alternative Proof of Identification: Linear Case (Theorem 2.6)

We also provide an alternative proof directly from the assumptions. It is more transparent and relies explicitly on properties of linear projections, but is not readily generalizable to the non-parametric setting.

Proof. Consider $\mathbb{E}^*[H|P, Z]$, that is, the linear projection of H on P and Z

$$H = \delta_0 + \delta_p P + \delta_z Z + \mu \tag{B.1}$$

where, by definition, the projection residual satisfies

$$\mathbb{E}[Z\mu] = 0, \quad \mathbb{E}[P\mu] = 0, \quad \mathbb{E}[\mu] = 0. \tag{B.2}$$

Now, under the linear single index assumption, we can substitute in the price function

$$\begin{aligned} H &= \delta_0 + \delta_p P + \delta_z Z + \mu && \text{(B.3)} \\ &= \delta_0 + \delta_p(\beta_0 + \beta_z Z + \Phi) + \delta_z Z + \mu && \text{(B.4)} \\ &= (\delta_0 + \delta_p\beta_0) + \delta_p\Phi + (\delta_z + \delta_p\beta_z)Z + \mu, && \text{(B.5)} \end{aligned}$$

where the projection residual also satisfies

$$\begin{aligned} \mathbb{E}[P\mu] &= \mathbb{E}[(\beta_0 + \beta_z Z + \Phi)\mu] && \text{(B.6)} \\ &= \beta_0\mathbb{E}[\mu] + \beta_z\mathbb{E}[Z\mu] + \mathbb{E}[\Phi\mu] && \text{(B.7)} \\ &= \mathbb{E}[\Phi\mu] = 0 && \text{(B.8)} \end{aligned}$$

Importantly, recognize that Eq (B.5) *is* the linear projection of H on Z and Φ because the term μ satisfies $\mathbb{E}[\Phi\mu] = 0$ by Eq (B.8), $\mathbb{E}[Z\mu] = 0$ and $\mathbb{E}[\mu] = 0$ by Eq (B.2). Thus, we have

$$\mathbb{E}^*[H|Z, \Phi] = (\delta_0 + \delta_p\beta_0) + \delta_p\Phi + (\delta_z + \delta_p\beta_z)Z. \tag{B.9}$$

In the meantime, because H satisfies the conditional orthogonality condition $cov(H, Z^{\perp\Phi}) = 0$, it means the linear projection of H on Z and Φ has a projection coefficient of zero in front of Z, or equivalently, $\mathbb{E}^*[H|Z, \Phi] = \mathbb{E}^*[H|\Phi]$. Hence, we have $\delta_z + \delta_p\beta_z = 0$. Lastly, the relevance

condition $cov(H, \Phi) \neq 0$ ensures that the projection coefficient of H on Φ is non-zero

$$\delta_p = \frac{cov(H, \Phi)}{var(\Phi)} \neq 0. \tag{B.10}$$

Therefore, the implicit price of amenity β_z is identified by the negative ratio of the two linear projection coefficients $\beta_z = -\delta_z/\delta_p$.

□

B.2 Proof of Asymptotics for Ratio-of-Coefficients Estimator (Theorem 2.8)

Proof. (Consistency) First, under the same regularity conditions in Theorem 2.6, the OLS estimator is a consistent estimator of the linear projection coefficients from H on Z and P, namely,

$$\hat{\delta_z} \xrightarrow{p} \delta_z, \ \hat{\delta_p} \xrightarrow{p} \delta_p. \tag{B.11}$$

Next, denote $\delta = (\delta_0, \delta_z, \delta_p)$. The mapping $g(\delta) = -\delta_z/\delta_p$ is continuous at δ as long as $\delta_p \neq 0$. Eq (B.10) from the proof for Theorem 2.6 has shown that the relevance condition $cov(H, \Phi) \neq 0$ implies $\delta_p \neq 0$. Therefore, consistency follows from the continuous mapping theorem.

(Asymptotic Normality) Additional regularity conditions $\mathbb{E}[H^4] < \infty$, $\mathbb{E}\|X\|^4 < \infty$, where $X = (Z, P, 1)'$, guarantee that the first-stage OLS estimator $\hat{\delta}$ is asymptotically normal with a variance-covariance matrix denoted by Σ. Again, since the relevance condition implies $\delta_p \neq 0$, the mapping $g(\delta) = -\delta_z/\delta_p$ is continuously differentiable in a neighborhood of δ. Therefore, the Delta method applies, where $G = \nabla g(\delta) = (0, -1/\delta_p, -\delta_z/\delta_p^2)$'. □

B.3 Derivation for Anderson-Rubin Confidence Interval

The following derivation for the ARCI for anti-IV is adapted from the derivation for conventional IV in Andrews, Stock, and Sun (2019). Specifically, let $h(\delta;\beta) = \delta_z + \beta\delta_p$, and let β_z denote the true coefficient. Then, the null hypothesis is transformed to $H_0 : \delta_z + \beta\delta_p = 0$. Since $\nabla h = (1, \beta)'$, the mapping $h(\cdot)$ is continuously differentiable everywhere regardless of the true δ_p, allowing us to apply the Delta method and obtain the following limit distribution

$$\hat{\delta}_z + \beta\hat{\delta}_p \xrightarrow{d} \mathcal{N}(0, \Omega(\beta)), \quad \text{where} \quad \Omega(\beta) = \Sigma_{zz} + 2\beta\Sigma_{zp} + \beta^2\Sigma_{pp} \tag{B.12}$$

where $\Sigma_{zz}, \Sigma_{zp}, \Sigma_{pp}$ refer to the specific covariance terms in the asymptotic variance-covariance matrix Σ of the OLS estimator $\hat{\delta}$. Then, redefine $h(\beta) = h(\hat{\delta};\beta)$, the Anderson-Rubin statistic follows a χ^2-distribution

$$AR(\beta) = \frac{h^2(\beta)}{\Omega(\beta)} \xrightarrow{d} \chi^2. \tag{B.13}$$

Correspondingly, the Anderson-Rubin confidence interval is $ARCI(\alpha) = \{\beta : AR(\beta) < \chi^2_{1-\alpha}\}$, where $\chi^2_{1-\alpha}$ denotes the $1-\alpha$ quantile of the χ^2 distribution. Therefore, regardless of the strength of the anti-instrument, as $n \to \infty$, the level $1-\alpha$ confidence interval can be solved analytically as the solution to the quadratic equation $h^2(\beta) = \chi^2_{1-\alpha}\Omega(\beta)$, yielding

$$ARCI(\alpha) = \left(-b - \frac{\sqrt{b^2 - 4ac}}{2a}, -b + \frac{\sqrt{b^2 - 4ac}}{2a}\right) \tag{B.14}$$

$$a = \delta_p^2 - \chi^2_{1-\alpha}\Sigma_{pp}, \quad b = 2(\delta_z\delta_p - \chi^2_{1-\alpha}\Sigma_{pz}), \quad c = \delta_z^2 - \chi^2_{1-\alpha}\Sigma_{zz}. \tag{B.15}$$

Note that the closed form solution for ARCI in Eq (B.14) is the asymptotic approximation as $n \to \infty$, but importantly, is robust to weak identification.

In our package *aivreg*, we report the 95% ARCI. In addition, since ARCI is typically not centered, we also report a pseudo-standard error by dividing the side of the ARCI closest to zero by 1.96 for easy comparison with conventional t-stats.

B.4 Proof of Identification: 2SLS

Theorem B.1. *(Identification) Under the assumptions in Theorem 2.6, the 2SLS coefficient is point-identified and identical to the ratio-of-coefficient*

$$\theta_{2SLS} = -\frac{\delta_z}{\delta_p} = \beta_z. \tag{B.16}$$

Proof. First, the predicted H based on the first-stage OLS regression of H on P and Z are

$$\tilde{H} = \mathbb{E}^*[H|P, Z] = \delta_0 + \delta_p P + \delta_z Z. \tag{B.17}$$

Then, the second-stage linear projection can be written as

$$\begin{aligned} P &= \phi_0 + \phi_z Z + \phi_{\tilde{h}} \tilde{H} + \omega && \text{(B.18)} \\ &= \phi_0 + \phi_z Z + \phi_{\tilde{h}}(\delta_0 + \delta_z Z + \delta_p P) + \omega && \text{(B.19)} \\ &= (\phi_0 + \phi_{\tilde{h}}\delta_0) + (\phi_z + \phi_{\tilde{h}}\delta_z)Z + \phi_{\tilde{h}}\delta_p P + \omega. && \text{(B.20)} \end{aligned}$$

Notice that P appears in both RHS and LHS, meaning the resulting linear projection has $\omega = 0$. Thus, with relevance guaranteeing that $\delta_p \neq 0$, we have

$$\phi_{\tilde{h}}\delta_p = 1 \;\Rightarrow \phi_{\tilde{h}} = \frac{1}{\delta_p} \tag{B.21}$$

$$\phi_0 + \phi_{\tilde{h}}\delta_0 = 0 \;\Rightarrow \phi_0 = -\phi_{\tilde{h}}\delta_0 = -\frac{\delta_0}{\delta_p} \tag{B.22}$$

$$\phi_z + \phi_{\tilde{h}}\delta_z = 0 \;\Rightarrow \phi_z = -\phi_{\tilde{h}}\delta_z = -\frac{\delta_z}{\delta_p}. \tag{B.23}$$

□

B.5 Equal Selection and Conditional Independence

Consider $\Phi = \lambda H + \Phi^u$ where λ represents the projection of Φ on H. Then, equal selection of H among unobservable controls can be thought of as $\mathbb{E}^*[Z|\lambda H, \Phi^u] = \psi_h \lambda H + \psi_u \Phi^u$, where $\psi_h = \psi_u = \psi$. Hence, the projection $\mathbb{E}^*[Z|H, \Phi] = \mathbb{E}^*[Z|\lambda H, \Phi^u] = \psi(\lambda H + \Phi^u) = \mathbb{E}^*[Z|\Phi]$, which means that H satisfies conditional orthogonality.

C GMM

In this appendix section, we derive all the relevant technical materials related to GMM mentioned in Section 2.2.5.

C.1 Method-of-Moment Estimator

As a stepping stone for developing GMM, we first consider a method-of-moments estimator for the just-identified case.

Definition C.1. (Method-of-Moments Estimator) Consider observations (H_i, Z_i, P_i) where $i = 1, \ldots, n$. Denote $\theta = (\beta_z, \theta_h, \theta_0)'$. Then, the moment conditions are

$$g_i(\theta) = \begin{bmatrix} (P_i - \beta_z Z_i - \theta_h H_i - \theta_0) Z_i \\ (P_i - \beta_z Z_i - \theta_h H_i - \theta_0) P_i \\ (P_i - \beta_z Z_i - \theta_h H_i - \theta_0) \end{bmatrix} \tag{C.1}$$

or, equivalently, let

$$X_i = \begin{pmatrix} Z_i \\ P_i \\ 1 \end{pmatrix}, \quad T_i = \begin{pmatrix} Z_i \\ H_i \\ 1 \end{pmatrix}, \quad g_i(\theta) = X_i(P_i - T_i'\theta). \tag{C.2}$$

Then, the method-of-moments estimator is

$$\hat{\theta}_{\text{MME}} = \left(\sum X_i T_i'\right)^{-1} \left(\sum X_i P_i\right). \tag{C.3}$$

To see where the first moment condition in Eq (C.1) comes from, we consider the hypothetical linear projection $\mathbb{E}^*[H|\Phi] = \gamma_0 + \gamma_\phi \Phi$, where γ denote the projection coefficients. Then, we can express conditional orthogonality as $\mathbb{E}^*[H|\Phi] = \mathbb{E}^*[H|Z, \Phi]$, which means that the projection residual must satisfy

$$\begin{aligned} 0 &= \mathbb{E}[(H - \gamma_0 - \gamma_\phi \Phi) Z] && \text{(C.4)} \\ &= \mathbb{E}[(H - \gamma_0 - \gamma_\phi (P - \beta_0 - \beta_z Z)) Z] && \text{(C.5)} \\ &= \mathbb{E}\Big[- \gamma_\phi \Big(P - \beta_z Z - \frac{1}{\gamma_\phi} H - \Big(\beta_0 - \frac{\gamma_0}{\gamma_\phi} \Big) \Big) Z \Big]. && \text{(C.6)} \end{aligned}$$

The other two moment conditions in Eq (C.1) follow from the requirement that the projection residual is uncorrelated with Φ, and consequently P due to invertibility, and has mean zero.

Lemma C.2. *(MME = Ratio-of-Coefficients) Under the assumptions in Theorem 2.6,* θ_{MME}

is point identified as

$$\theta_{MME} = \mathbb{E}[XT']^{-1}\mathbb{E}[XP], \tag{C.7}$$

and its first element is the ratio-of-coefficients:

$$\theta_{MME,1} = \beta_z = -\frac{\delta_z}{\delta_p}, \tag{C.8}$$

where δ_p and δ_z are the linear projection coefficients of H on P and Z.

Proof. See Supplemental Appendix C.4.1. □

Lemma C.2 implies that the MME estimator is equivalent to the ratio-of-coefficients estimator.

C.2 GMM Estimator

We now generalize MME to GMM to allow for more than one instrument by stacking the three moment conditions in Eq (2.24) for each anti-instrument.

Definition C.3. (GMM Estimator) Consider observations $(H_i^1, \ldots, H_i^L, Z_i, P_i)$ where $i = 1, \ldots, n$ and $l = 1, \ldots, L$ indexes the anti-instruments. Let $\theta = (\beta_z, \theta_h^1, \theta_0^1, \ldots, \theta_h^l, \theta_0^l, \ldots, \theta_h^L, \theta_0^L)'$, and

$$X_i = \begin{pmatrix} Z_i & 0 & 0 & & & \\ 0 & P_i & 0 & & & \\ 0 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & & & \\ & & & & & \end{pmatrix}, \; T_i = \begin{pmatrix} Z_i & Z_i & Z_i & \ldots & Z_i & Z_i & Z_i \\ H_{i,1} & H_{i,1} & H_{i,1} & \ldots & 0 & 0 & 0 \\ 1 & 1 & 1 & \ldots & 0 & 0 & 0 \\ & & & \ldots & & & \\ 0 & 0 & 0 & \ldots & H_{i,L} & H_{i,L} & H_{i,L} \\ 0 & 0 & 0 & \ldots & 1 & 1 & 1 \end{pmatrix}, \; P_i\mathbf{1} = \begin{pmatrix} P_i \\ P_i \\ P_i \\ \vdots \\ P_i \\ P_i \\ P_i \end{pmatrix}$$

where the dimension of X_i is $3L \times 3L$, the dimension of T_i is $(2L+1) \times 3L$, the dimension of θ is $(2L+1) \times 1$, and with a slight abuse of notation, let P_i denote the length $2L+1$ vector $P_i\mathbf{1}$. The moment conditions can be written as

$$g_i(\theta) = X_i(P_i - T_i'\theta). \tag{C.9}$$

With a positive definite weighting matrix W, the GMM-estimator is

$$\hat{\theta}_{\text{GMM}} = \left(\left(\sum X_i T_i'\right)' W \left(\sum X_i T_i'\right)\right)^{-1} \left(\left(\sum X_i T_i'\right)' W \left(\sum X_i P_i\right)\right). \tag{C.10}$$

Theorem C.4. *(Identification for GMM) Under the assumptions in Theorem 2.6, where all anti-instruments H_l, $l = 1, \ldots, L$, satisfy the relevance and conditional orthogonality conditions, the GMM coefficient is point-identified as*

$$\theta_{GMM} = \left(\left(\mathbb{E}[XT'] \right)' W \left(\mathbb{E}[XT'] \right) \right)^{-1} \left(\left(\mathbb{E}[XT'] \right)' W \left(\mathbb{E}[XP] \right) \right). \tag{C.11}$$

Proof. See Supplemental Appendix C.4.2. □

Given the GMM coefficient is point identified, under mild regularity conditions, consistency and asymptotic normality of the efficient GMM estimator follow, namely,

$$\hat{\beta}_{\text{GMM}} \equiv \hat{\theta}_{\text{GMM},1} \xrightarrow{p} \beta_z, \tag{C.12}$$

$$\sqrt{n}(\hat{\theta}_{\text{GMM}} - \theta_{\text{GMM}}) \xrightarrow{d} \mathcal{N}(0, V_\theta), \quad V_\theta = (Q'\Omega^{-1}Q)^{-1}, \tag{C.13}$$

where $Q = \mathbb{E}[XT']$, $\Omega = \mathbb{E}[(Xe)(Xe)']$, with $e = P - T'\theta$, and $W = \Omega^{-1}$ is the efficient weighting matrix.

Therefore, the GMM estimator generalizes the ratio-of-coefficients estimator in the just-identified case, and with an appropriate weighting matrix, provides efficient estimation in the over-identified case. Additionally, the GMM formulation also allows us to use its objective,

$$J(\hat{\theta}_{\text{GMM}}) = \sqrt{n}\bar{g}_n(\hat{\theta}_{\text{GMM}})' \, \hat{\Sigma}^{-1} \, \sqrt{n}\bar{g}_n(\hat{\theta}_{\text{GMM}}) \tag{C.14}$$

as a test statistic for the corresponding over-identification test.

C.3 2SLS as GMM

There is also a "2SLS" implementation of the anti-IV approach, which directly regresses price P on amenity Z while controlling for an estimator of the predicted anti-IV,

$$\hat{H}_i^l = \hat{\delta}_0^l + \hat{\delta}_p^l P_i + \hat{\delta}_z^l Z_i, \tag{C.15}$$

where $\hat{\delta}^l$ are the "first-stage" OLS regression coefficients obtained by regressing H^l on P and Z. The "2SLS" estimator is obtained by an OLS regression of P on Z and $\hat{H}^1, \ldots, \hat{H}^l$ as the "second-stage":

$$P_i = \hat{\phi}_0 + \hat{\phi}_z Z_i + \hat{\phi}_{\hat{h}_1} \hat{H}_i^1 + \ldots + \hat{\phi}_{\hat{h}_L} \hat{H}_i^L + \omega_i, \tag{C.16}$$

$$\text{where } \sum \omega_i Z_i = 0, \quad \forall l : \sum \omega_i \hat{H}_i^l = 0, \quad \sum \omega_i = 0, \tag{C.17}$$

and let $\hat{\theta}_{\text{2SLS}} \equiv \hat{\phi}_z$.

The "2SLS" estimator can be represented as a GMM estimator with an alternative weighting matrix $W = (\mathbb{E}[XX'])^{-1}$, namely, the optimal weighting matrix under homoskedasticity. The 2SLS estimator of conventional IVs also has an analogous property, for instance, see Theorem 13.2 in Hansen (2022). We leave the derivations to the Supplemental Appendix C.4.3.

C.4 Additional Proofs for GMM

C.4.1 Proof of Identification: MME (Lemma C.2)

Proof. For point identification, it suffices to show that $\mathbb{E}[XT']$ is invertible. We compute its determinant explicitly as follows

$$\mathbb{E}[XT'] = \mathbb{E}\left[\begin{pmatrix} Z \\ P \\ 1 \end{pmatrix} \begin{pmatrix} Z & H & 1 \end{pmatrix}\right] = \mathbb{E}\begin{bmatrix} ZZ & ZH & Z \\ PZ & PH & P \\ Z & H & 1 \end{bmatrix} = \mathbb{E}\begin{bmatrix} \sigma_{ZZ} + \mu_Z^2 & \sigma_{ZH} + \mu_Z\mu_H & \mu_Z \\ \sigma_{ZP} + \mu_Z\mu_p & \sigma_{PH} + \mu_P\mu_H & \mu_P \\ \mu_Z & \mu_H & 1 \end{bmatrix}, \tag{C.18}$$

where μ and σ denote the corresponding means and covariances for Z, H, P. Then, its determinant becomes

$$\det\ \mathbb{E}[XT'] = \det\ \mathbb{E}\begin{bmatrix} \sigma_{ZZ} & \sigma_{ZH} & 0 \\ \sigma_{ZP} & \sigma_{PH} & 0 \\ \mu_Z & \mu_H & 1 \end{bmatrix} = \sigma_{ZZ}\sigma_{PH} - \sigma_{ZH}\sigma_{ZP}. \tag{C.19}$$

Note that we can drop all the mean terms by subtracting out a linear combinations of the last row.

Consider the linear projection from H to Z and P where $\mathbb{E}^*[H|Z,P] = \delta_z Z + \delta_p P + \delta_0$, then

$$\delta_p = \frac{cov(H, P^{\perp Z})}{var(P^{\perp Z})}, \quad cov(H, P^{\perp Z}) = cov\left(H, P - \frac{cov(P,Z)}{var(Z)}Z\right) = \left(\sigma_{PH}\sigma_{ZZ} - \sigma_{ZP}\sigma_{ZH}\right)/\sigma_{ZZ} \tag{C.20}$$

$$var(P^{\perp Z}) = var\left(P - \frac{cov(P,Z)}{var(Z)}Z\right) = \left(\sigma_{ZZ}\sigma_{PP} - \sigma_{ZP}^2\right)/\sigma_{ZZ}. \tag{C.21}$$

The denominator $(\sigma_{ZZ}\sigma_{PP} - \sigma_{ZP}^2) \neq 0$ because of the regularity condition $\mathbb{E}[XX']$ being positive definite. Recall that the proof of Theorem 2.6 shows that the relevance condition implies that $\delta_p \neq 0$, thus implying that the numerator $\det\ \mathbb{E}[XT'] = \sigma_{ZZ}\sigma_{PH} - \sigma_{ZH}\sigma_{ZP} \neq 0$. Therefore, θ_{MME} is point identified.

Next, to obtain the explicit formula for the first element, we compute the determinant of the

auxiliary matrix where we replace the first column with $\mathbb{E}[XP]$ as follows

$$\det \mathbb{E}[XT']_{(1)} = \det \mathbb{E}\begin{bmatrix} ZP & ZH & Z \\ PP & PH & P \\ P & H & 1 \end{bmatrix} = \sigma_{ZP}\sigma_{PH} - \sigma_{PP}\sigma_{ZH}, \tag{C.22}$$

Then, Cramer's Rule implies that the first element of the solution to $\mathbb{E}[XT']\theta_{\text{MME}} = \mathbb{E}[XP]$ is

$$\theta_{\text{MME},1} = \frac{\sigma_{ZP}\sigma_{PH} - \sigma_{PP}\sigma_{ZH}}{\sigma_{ZZ}\sigma_{PH} - \sigma_{ZH}\sigma_{ZP}}. \tag{C.23}$$

In terms of the linear projection from H to Z and P, we also have

$$\delta_z = \frac{cov(H, Z^{\perp P})}{var(Z^{\perp P})}, \quad cov(H, Z^{\perp P}) = cov\left(H, Z - \frac{cov(P,Z)}{var(P)}P\right) = \left(\sigma_{ZH}\sigma_{PP} - \sigma_{ZP}\sigma_{PH}\right)/\sigma_{PP} \tag{C.24}$$

$$var(Z^{\perp P}) = var\left(Z - \frac{cov(P,Z)}{var(P)}P\right) = \left(\sigma_{ZZ}\sigma_{PP} - \sigma_{ZP}^2\right)/\sigma_{PP} \tag{C.25}$$

Together, we have

$$\theta_{\text{MME},1} = \frac{\sigma_{ZP}\sigma_{PH} - \sigma_{PP}\sigma_{ZH}}{\sigma_{ZZ}\sigma_{PH} - \sigma_{ZH}\sigma_{ZP}} = \frac{cov(H, P^{\perp Z})/var(P^{\perp Z})}{cov(H, Z^{\perp P})/var(P^{\perp Z})} = -\frac{\delta_z}{\delta_p} = \beta_z. \tag{C.26}$$

□

C.4.2 Proof of Identification: GMM (Theorem C.4)

Proof. It suffices to show that $Q \equiv \mathbb{E}[XT']$ has a full column rank of $2L+1$, so that $\left(\mathbb{E}[XT']\right)' W \left(\mathbb{E}[XT']\right)$ is invertible for any positive definite W. We can write

$$\mathbb{E}[XT'] = \mathbb{E}\begin{bmatrix} Q^{(1)} \\ \vdots \\ Q^{(L)} \end{bmatrix}, \quad Q^{(l)} = \begin{bmatrix} ZZ & ZH_l & Z \\ PZ & PH_l & P \\ Z & H_l & 1 \end{bmatrix} = \left[\begin{pmatrix} Z \\ P \\ 1 \end{pmatrix}\begin{pmatrix} Z & H_l & 1 \end{pmatrix}\right] \tag{C.27}$$

where the 3×3 block sits in columns $(1, 2l, 2l+1)$ and zero elsewhere.

If Q does not have full column rank, then there exists a non-zero vector $\theta = (\theta_z, \theta_h^1, \theta_0^1, \ldots, \theta_h^L, \theta_0^L) \neq 0$ such that $Q\theta = 0$. Then, based on the block structure of Q, there must exist some block l with $\theta_l = (\theta_z, \theta_h^l, \theta_0^l) \neq 0$ such that $Q^{(l)}\theta_l = 0$. However, since H_l is relevant, based on the proof in Theorem C.2 we know that $Q^{(l)}$ is invertible, so $\theta_l = \left(Q^l\right)^{-1} 0 = 0$, constituting a contradiction. Therefore, $Q = \mathbb{E}[XT']$ must have full column rank. Hence, when all anti-instruments H_l are relevant, the GMM coefficient is point identified. □

C.4.3 "2SLS" Equivalence to GMM

Lemma C.5. *The "2SLS" estimator is equivalent to the GMM estimator with the weighting matrix estimated from $W = V^{-1}$, $V = \mathbb{E}[XX']$, which is the efficient weighting matrix under homoskedasticity.*

Proof. With $Q = \mathbb{E}[XT']$, we can write the GMM coefficients with weighting V^{-1} as

$$Q'V^{-1}Q\,\theta_{\text{GMM}} = Q'V^{-1}\mathbb{E}[XP]. \tag{C.28}$$

Now, consider the linear projection of T on X

$$\tilde{T} = \left(\mathbb{E}[XX]^{-1}\mathbb{E}[XT']\right)' X = Q'V^{-1}X \tag{C.29}$$

Then, we have

$$\mathbb{E}[\tilde{T}\tilde{T}'] = \mathbb{E}[(Q'V^{-1}X)(Q'V^{-1}X)'] = \mathbb{E}[Q'V^{-1}XX'V^{-1}Q] = Q'V^{-1}Q \tag{C.30}$$

$$\mathbb{E}[\tilde{T}P] = \mathbb{E}[Q'V^{-1}XP] = Q'V^{-1}\mathbb{E}[XP] \tag{C.31}$$

which allows us to redefine the GMM coefficients with a weighting matrix of V^{-1} as the projection coefficients of P on $\tilde{T}$

$$\mathbb{E}[\tilde{T}\tilde{T}']\,\theta_{\text{GMM}} = \mathbb{E}[\tilde{T}P], \tag{C.32}$$

which is precisely the "2SLS" estimator when projecting each H_l on $(Z, P, 1)$ and use them in a "second-stage" regression.

□

D Generalization to Non-Parametric Identification

First, the following lemma provides the foundation for non-parametric identification.

Lemma D.1. *Consider a market that satisfies the single index assumption (Assumption 2.10). If an anti-instrumental variable H satisfies conditional independence (Assumption 2.13), then*

$$\tilde{H}(P, Z) = \tilde{H}(\Phi(P, Z)). \tag{D.1}$$

Proof.

$$\begin{aligned}
&\tilde{H}(p, z) && && \text{(D.2)}\\
&= \mathbb{E}[H|P = p, Z = z] && \text{by definition} && \text{(D.3)}\\
&= \mathbb{E}[H|P = p, Z = z, \Phi = \phi] && \text{by the invertibility } \phi = \Phi(p, z) \text{ due to } \textit{single index} && \text{(D.4)}\\
&= \mathbb{E}[H|Z = z, \Phi = \phi] && \text{by the existence of the price function } P(z, \phi) && \text{(D.5)}\\
&= \mathbb{E}[H|\Phi = \phi] = \tilde{H}(\phi). && \text{by conditional independence} && \text{(D.6)}
\end{aligned}$$

□

Now, we provide the proof for the main theorem (Theorem 2.14).

Proof. Lemma D.1 implies

$$\tilde{H}(P(Z, \Phi), Z) = \tilde{H}(\Phi) \tag{D.7}$$

Take the total derivative of the LHS of Equation (D.7) at (z, ϕ)

$$\frac{d}{dZ}\Big(\tilde{H}(P(Z, \Phi), Z)\Big)\Big|_{(z,\phi)} = \frac{\partial \tilde{H}(P, Z)}{\partial P}\Big|_{(p,z)} \frac{\partial P(Z, \Phi)}{\partial Z}\Big|_{(z,\phi)} + \frac{\partial \tilde{H}(P, Z)}{\partial Z}\Big|_{(p,z)} \tag{D.8}$$

Correspondingly, take the total derivative of the RHS of Equation (D.7)

$$\frac{d}{dZ}\Big(\tilde{H}(\Phi)\Big)\Big|_{\phi} = 0. \tag{D.9}$$

Taken together, we have

$$\frac{\partial P(Z, \Phi)}{\partial Z}\Big|_{(z,\phi)} = -\frac{\partial \tilde{H}(P, Z)}{\partial Z} \Big/ \frac{\partial \tilde{H}(P, Z)}{\partial P}\Big|_{(p,z)} \tag{D.10}$$

where the denominator

$$\left.\frac{\partial \tilde{H}(P,Z)}{\partial P}\right|_{(p,z)} = \left.\frac{\partial \tilde{H}(\Phi)}{\partial \Phi}\right|_{\phi} \times \left.\frac{\partial \Phi(P,Z)}{\partial P}\right|_{(p,z)} \neq 0. \tag{D.11}$$

The last statement is based on the relevance condition, ensuring $\partial \tilde{H}(\Phi)/\partial \Phi|_{\phi} \neq 0$. Then, the single index assumption ensures that $\partial P(Z,\Phi)/\partial \Phi \neq 0$, which, under continuous differentiability, implies $\partial \Phi(P,Z)/\partial P \neq 0$. □

E "Missing" Amenities

When there are missing amenities, our model becomes mis-specified, and the estimated amenity price by our anti-IV approach may differ from the true implicit price.

First, one approach to address this issue could be to require a candidate anti-IV to be independent from the focal amenity but conditioning on the unobserved quality that includes the impact of the missing amenity. Then, the usual approach of anti-IV identification and estimation follows. This approach most likely applies in scenarios where the "missing" amenity can be plausibly viewed as part of the unobserved quality.

However, if the missing amenity cannot be folded into the unobserved quality term without violating the requisite conditional independence, we derive a "bias" formula for the anti-IV approach, analogous to the OLS bias formula, which is also biased in the presence of missing amenities. The formula highlights that the bias becomes more pronounced if the amenity is important for prices, and if it is also heavily correlated with the focal amenity even after controlling for the adjusted quality.

Thus, with missing amenities, our anti-IV approach estimates the price a home-buyer would pay to obtain the additional amenity, but also factoring in the remaining correlation between the observed amenity and unobserved amenities after controlling for adjusted quality. While this may differ from the impact on home prices due to an exogenous and isolated change in the same amenity, as found in quasi-experiments, we believe it could still be interpretable as it is pertinent for assessing housing inequality, for instance, the actual amount home-buyers would need to pay to attain a certain amenity level.

Additionally, unlike the OLS methods, we also show that the missing amenity bias of the anti-IV estimator remains stable even when quality may be poorly measured. This suggests that our anti-IV method can offer a more reliable price estimate compared to traditional OLS methods, even when there are "missing" amenities.

E.1 Set-Up

Consider a market where prices are determined by two amenities Z_1, Z_2 and quality Φ. We focus on the linear case where the pricing equation is

$$P = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \Phi \tag{E.1}$$

where $\beta_1, \beta_2 \neq 0$. Suppose we only observe Z_1, then we have a mis-specified model with an adjusted quality $\tilde{\Phi}$

$$P = \beta_0 + \beta_1 Z_1 + \tilde{\Phi}, \quad \text{where } \tilde{\Phi} = \Phi + \beta_2 Z_2. \tag{E.2}$$

Before we get into the bias formula, we want to first note that the "missing" amenity issue

may not be problematic if we can find an appropriate anti-IV that satisfied the identifying restrictions with respect to the adjusted quality $\tilde{\Phi}$, namely,

$$cov(H, \tilde{\Phi}) = 0 \quad \text{and} \quad cov(H, Z_1^{\perp \tilde{\Phi}}) = 0 \tag{E.3}$$

then the usual anti-IV approach applies and there is no bias. While this may seem mechanical, this approach is likely applicable when the "missing" amenity can be readily thought as part of the unobserved quality and the candidate anti-IV remains plausibly uncorrelated with the focal amenity when conditioning on that adjusted quality.

Now, let us consider an anti-IV variable H that satisfies conditional orthogonality but only with respect to the true quality Φ, namely,

$$cov(H, Z_1^{\perp \Phi}) = 0. \tag{E.4}$$

Then, the amenity price would be identified using the ratio of coefficients formula if all amenities were observed

$$\beta_1 = -\frac{\delta_1}{\delta_p}, \tag{E.5}$$

where the coefficients are based on the linear projection

$$\mathbb{E}^*[H|Z_1, Z_2, P] = \delta_0 + \delta_1 Z_1 + \delta_2 Z_2 + \delta_p P. \tag{E.6}$$

E.2 Formula for Missing Amenity Bias - OLS

For comparison, we first derive the bias using OLS while controlling for H, an imperfect measure for quality Φ,

$$\mathbb{E}^*[P|Z_1, H] = \beta_0^{\text{ols}} + \beta_1^{\text{ols}} Z_1 + \beta_H^{\text{ols}} H.$$

We derive the OLS bias formula when quality is imperfectly controlled for

$$\beta_1^{\text{ols}} = \frac{cov(P, Z_1^{\perp H})}{var(Z_1^{\perp H})} \tag{E.7}$$

$$= \frac{cov(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \Phi, Z_1^{\perp H})}{var(Z_1^{\perp H})} \tag{E.8}$$

$$= \beta_1 + \beta_2 \frac{cov(Z_2, Z_1^{\perp H})}{var(Z_1^{\perp H})} + \frac{cov(\Phi, Z_1^{\perp H})}{var(Z_1^{\perp H})} \tag{E.9}$$

Note that the first bias component $cov(Z_2, Z_1^{\perp H})$ is driven by the correlation between observed and unobserved amenity. The second bias component $cov(\Phi, Z_1^{\perp H})$ is caused by the fact that H is an imperfect control for Φ.

Thus, in the ideal case for OLS when H perfectly measures Φ (e.g., $H = \Phi$), OLS is biased due the missing amenity term only, namely, $cov(Z_2, Z_1^{\perp\Phi})$. However, in the more typical case when H is just relevant for Φ, OLS bias come from both the missing amenity term $cov(Z_2, Z_1^{\perp H})$ and the missing quality term $cov(\Phi, Z_1^{\perp H})$.

As an illustration, suppose Z_1 represents "clean air" and Z_2 represents "clear roads", and to the extent that they are positively correlated when controlling for housing quality Φ, then the OLS estimator provides a positively biased estimator. To interpret the "missing amenity bias" from the OLS estimate, it essentially means that a home-buyer cannot live in a neighborhood with clean air without also having to pay for the fact that these are less congested neighborhoods, and that is because the presence of these amenities in the market is positively correlated even when conditioning on quality.

E.3 Formula for Missing Amenity Bias - Anti-IV

Next, we derive the bias formula for the anti-IV estimator when there are missing amenities. Suppose Z_2 is missing, then, the anti-IV estimator becomes

$$\beta_1^{\text{aiv}} = -\frac{\gamma_1}{\gamma_p} \tag{E.10}$$

where the coefficients are based on the linear projection of H on P and Z_1 alone

$$\mathbb{E}^*[H|Z_1, P] = \gamma_0 + \gamma_1 Z_1 + \gamma_p P. \tag{E.11}$$

Let $\epsilon^{1,2,P}$ denote the projection residual from

$$\mathbb{E}^*[H|Z_1, Z_2, P] = \delta_0 + \delta_1 Z_1 + \delta_2 Z_2 + \delta_p P, \tag{E.12}$$

then, by definition, it would satisfy

$$cov(\epsilon^{1,2,P}, Z_1) = 0, \quad \text{and} \quad cov(\epsilon^{1,2,P}, P) = 0. \tag{E.13}$$

We can rewrite the numerator of the anti-IV estimator as follows:

$$\gamma_1 = \frac{cov(H, Z_1^{\perp P})}{var(Z_1^{\perp P})} \tag{E.14}$$

$$= \frac{cov(\delta_0 + \delta_1 Z_1 + \delta_2 Z_2 + \delta_p P + \epsilon^{1,2,P},\ Z_1^{\perp P})}{var(Z_1^{\perp P})} \tag{E.15}$$

$$= \delta_1 + \delta_2 \frac{cov(Z_2, Z_1^{\perp P})}{var(Z_1^{\perp P})} \tag{E.16}$$

Similarly, we can rewrite the denominator of the anti-IV estimator as follows:

$$\gamma_p = \frac{cov(H, P^{\perp Z_1})}{var(P^{\perp Z_1})} \tag{E.17}$$

$$= \frac{cov(\delta_0 + \delta_1 Z_1 + \delta_2 Z_2 + \delta_p P + \epsilon^{1,2,P},\ P^{\perp Z_1})}{var(P^{\perp Z_1})} \tag{E.18}$$

$$= \delta_p + \delta_2 \frac{cov(Z_2, P^{\perp Z_1})}{var(P^{\perp Z_1})}. \tag{E.19}$$

Note that the covariance term $cov(\epsilon^{1,2,P}, Z_1^{\perp P}) = 0$ and $cov(\epsilon^{1,2,P}, P^{\perp Z_1}) = 0$, because the residual term $\epsilon^{1,2,P}$ is by construction uncorrelated with P and Z_1 due to Eq (E.13).

Putting the two together, the anti-IV coefficient can be written as

$$\beta_1^{\text{aiv}} = -\frac{\gamma_1}{\gamma_p} = -\frac{\delta_1 + \delta_2 \times \psi_1}{\delta_p + \delta_2 \times \psi_p} \tag{E.20}$$

where ψ_1 and ψ_p is defined by the linear regression of Z_2 on P and Z_1

$$\mathbb{E}^*[Z_2|Z_1, P] = \psi_0 + \psi_1 Z_1 + \psi_p P. \tag{E.21}$$

We make several observations with respect to the bias formula Eq (E.20).

First, the degree of anti-IV bias is driven in part by the magnitude of δ_2, or, equivalently, $\beta_2 = -\delta_2/\delta_p$. In other words, if the missing amenity is not a big contributor to price, it is also not a big contributor to the bias.

Second, the degree of bias is also driven by how $-\psi_1/\psi_p$ departs from $-\delta_1/\delta_p = \beta_1$. In other words,

$$\text{if} - \frac{\psi_1}{\psi_p} = \beta_1, \quad \text{then} \quad \hat{\beta}_1^{\text{aiv}} = \beta_1. \tag{E.22}$$

Namely, the aiv coefficient is unbiased if $-\psi_1/\psi_p = \beta_1$. Note that the condition $-\psi_1/\psi_p = \beta_1$ can also be viewed as using the projection $\mathbb{E}^*[Z_2|Z_1, P]$ as the estimating first stage, which requires Z_2 to be a valid "anti-IV" in this setting, namely, satisfying the following conditional orthogonality condition

$$cov(Z_2, Z_1^{\perp \tilde{\Phi}}) = 0. \tag{E.23}$$

Thus, for the missing amenity to lead to no bias, we would require Z_2 to be uncorrelated with the focal amenity Z_1 once conditioning on the adjusted quality $\tilde{\Phi}$. Admittedly, the conditional orthogonality condition in Eq (E.23) may well be a knife-edge case, but it also clarifies the exact source of the bias.

Third, since we have an anti-IV variable H, unlike OLS, the anti-IV coefficient is *not* sensitive to how well the quality Φ itself is measured by H. Thus, the only source of bias for the anti-IV

approach is determined by the correlation of Z_2 and Z_1 conditioning on P. Therefore, even though neither the anti-IV method nor the OLS method is inherently designed to address the issue of missing amenities, the anti-IV approach remains robust to poorly measured quality even in the presence of missing amenities.

As an illustration, suppose again Z_1 represents "clean air" and Z_2 represents "clear roads". They could remain positively correlated even after controlling for the adjusted quality $\tilde{\Phi}$, which encapsulates both the unobserved quality Φ and a contribution from $\beta_2 Z_2$. Then, to interpret the "missing amenity bias" from the anti-IV estimate, it essentially means that a home-buyer cannot live in a neighborhood with clean air without also having to pay some for the less congested roads, and that is because the presence of these amenities in the market can remain positively correlated even when conditioning on the adjusted quality.

E.4 Simulation

We illustrate the property of both OLS and anti-IV when there is an unobserved amenity by running simulations using a DGP based on Cobb-Douglas preferences and competitive supply.

DGP

The data generating process is based on a Cobb-Douglas demand over two amenities Z_1 and Z_2.

$$U_i(Z_1, Z_2, \Phi, C) = Z_1^{\theta_i^1} Z_2^{\theta_i^2} \Phi^{\theta^\phi} C^{\theta_i^C} \tag{E.24}$$

where $\theta_i^1 + \theta_i^2 + \theta^\phi + \theta_i^C = 1$. The first amenity Z_1 is observed by the researcher, but Z_2 is not. The goal is to correctly estimate the price of the observed amenity β_1. Under this framework, maximizing utility with a budget constraint of η_i

$$\max_{Z_1, Z_2, \Phi, C} U_i(Z_1, Z_2, \Phi, C) \qquad \text{s.t.} \quad \beta_1 Z_1 + \beta_2 Z_2 + \Phi + C \leq \eta_i \tag{E.25}$$

results in the following optimal choice for amenity and quality,

$$Z_{i,1}^* = \frac{\theta_i^1}{\beta_1}\eta_i, \quad Z_{i,1}^* = \frac{\theta_i^2}{\beta_2}\eta_i, \quad \Phi_i^* = \theta_i^\phi \eta_i \tag{E.26}$$

where the expenditure share on quality expands proportionally with one's budget η_i and the expenditure share on amenity is further modulated by the strength of one's preference.

For the simulation, let the true price for the amenities be $(\beta_1, \beta_2) = (0.3, 0.3)$.

Households budget sets η_i are simulated as follows: The set of all possible values of household budgets η_i is $\mathcal{E} = \{10n | n = 1, 2, ..., 100\}$. For each value in $\mathcal{E}$, there are exactly 50,000 households with this amount of budget, so there are a total of 5,000,000 households in the simulation. We simulate under both the case when η_i is perfectly observed, and when η_i is observed with a

random noise as $h_i = 0.7\eta_i + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, 100)$ is the random noise.

The preference parameter for quality, θ^ϕ, is 0.5. The preference parameter for the amenities, θ_i^1 and θ_i^2, are drawn from a joint distribution, independently of η_i. Each has a uniform marginal distribution on $[0, 0.25)$, and the correlation between them is $corr(\theta_i^1, \theta_i^2) = \rho$. The correlation ρ is a key determinant of the direction of missing amenity bias, since it captures the co-movement of the demand for the two amenities as the expenditure on each amenity is driven by one's preferences. Therefore, we examine the results under various correlation ρ in our simulations.

Simulation Results

We simulate the data by varying the correlation between the preference for the observed and unobserved amenity to examine how the anti-IV and OLS results respond to the change in ρ. The results from the simulations are presented in Figure E.1. The x-axis represents the correlation between preference for amenities in each simulation, while the colored series depicts the estimated price for amenity one using different methods.

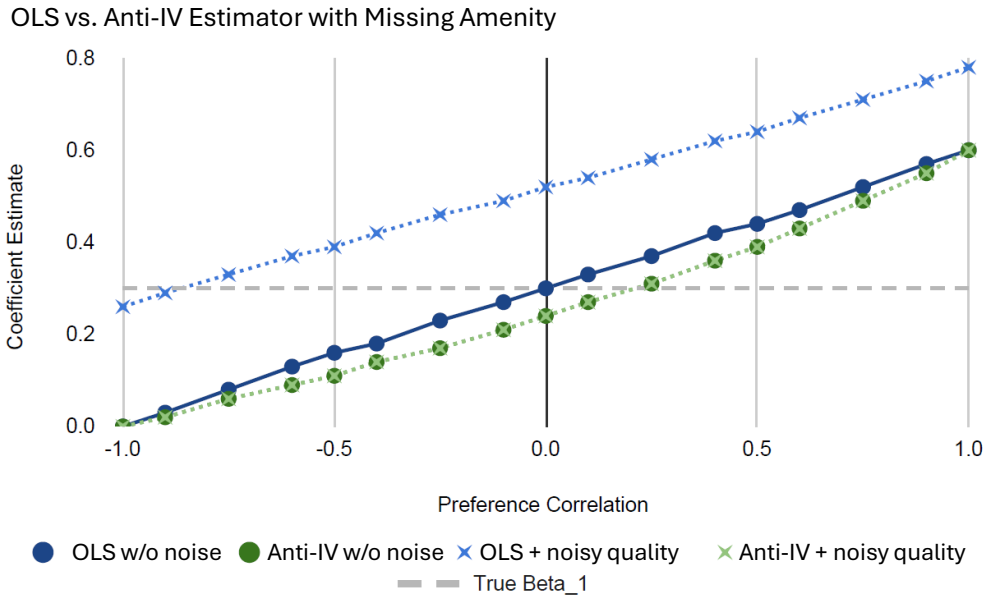
First, let us consider a scenario where all household budget sets, and therefore the amount of housing quality chosen, is perfectly observed. In this case, the solid blue and green lines indicate the estimated price for amenity using OLS and anti-IV, respectively. Notice that in the absence of correlation between preferences, the OLS method (solid blue) restores the original price β_1. The anti-IV method (solid green) also restores the correct price but at a different point when there is some positive correlation. In other words, the anti-IV can be unbiased even when Z_1 and Z_2 are somewhat positively correlated as long as Z_2 can remain uncorrelated with Z_1 controlling for $\tilde{\Phi}$ as in Eq (E.23). Moreover, these two lines intersect at both ends, where the preferences for the two amenities are perfectly correlated, $\rho = \pm 1$. In this case, as expected, the price of Z_2 cannot be separately identified from Z_1, and both methods are reduced to estimating the price of the single combined amenity.

Lastly, when the household budget sets are observed with noise, the OLS estimator yields a highly biased result (dotted blue line), while the anti-IV estimator remains unaffected and robust to components of unobserved quality (dotted green line). Furthermore, when we explore how sensitive our estimate is to the relationship between the anti-IV h and the budget set η, we simulate the data by holding the total variance of η fixed while varying the signal to total ratio of h.[29] Figure E.2 shows that as the signal-to-total ratio decreases (going from the right of the graph to the left), the bias from OLS with anti-IV control increases steadily, while

[29]To simulate this, households have a random draw of both anti-IV and noise, with $h_i \sim \mathcal{N}(0, 100)$ and $\epsilon_i \sim \mathcal{N}(0, 100)$, and the budget set is given as $\eta_i = \lambda h_i + (1 - \lambda)\epsilon_i$, where λ denotes varying levels of how informative h is.

the bias for the anti-IV method is unchanged across different levels of signal to total ratio.[30] This simulation result is consistent with our derivation where the missing amenity "bias" of the anti-IV estimator is robust to varying levels of signal-to-total ratio of the anti-IV variable, which is a property that OLS does not have.

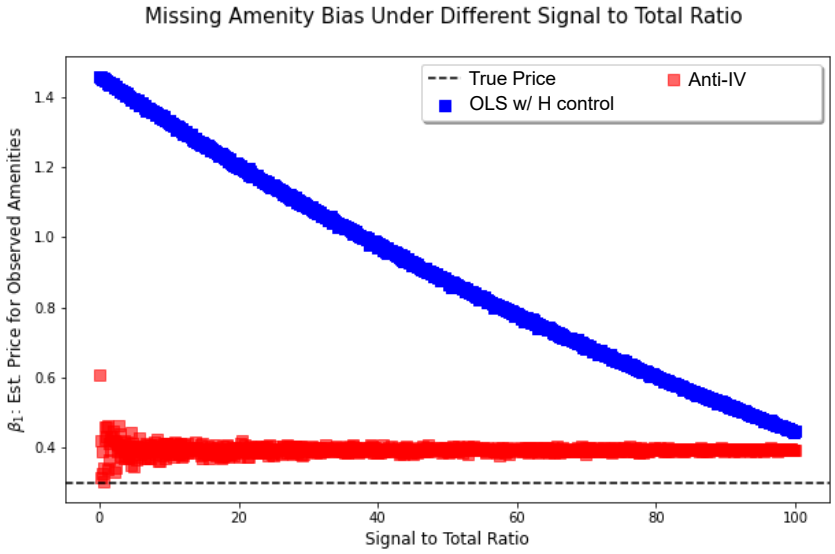
Figure E.1: Missing Amenities Bias



Notes: This figure compares the OLS and anti-IV estimator with a missing amenity under varying levels of preference correlations for the two underlying amenities. OLS estimates are shown in blue lines, and the anti-IV estimates are shown in green lines. Both methods can suffer from the missing amenity bias. However, in the presence of imperfectly measured quality, the bias in the OLS estimate (dotted blue line) is substantial, whereas the anti-IV estimator is not affected (dotted green line).

[30] When the signal-to-total ratio is very close to zero, the anti-IV method suffers from a "weak" anti-IV problem, analogous to the weak instrument problem of an IV estimator, where the denominator in the ratio-of-coefficients becomes close to zero.

Figure E.2: Missing Amenity Bias v.s Signal to Total Ratio



Notes: This figure shows that as the signal-to-total ratio decreases (going from the right of the graph to the left), the bias from OLS with anti-IV control increases steadily, while the bias for the anti-IV method is unchanged across different levels of signal-to-total ratio.