

Worker Utility as Hysteresis: A Preisach Model of Transaction Acceptance in Gig Labour Markets

PIOTR FRYDRYCH

`piotr.frydrych@pw.edu.pl`

Warsaw University of Technology

Roman Szewczyk

Polish Academy of Sciences

Research Article

Keywords: Preisach model, labor market hysteresis, worker utility, gig economy, latent threshold estimation, discrete choice

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9594317/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Worker Utility as Hysteresis: A Preisach Model of Transaction Acceptance in Gig Labour Markets

[Author Names — Blinded for Review]

[affiliation blinded for review]

Corresponding author: [email — blinded for review]

Abstract

Worker utility is not observed — only its consequence is. Each transaction on a gig platform produces a single bit: accepted or rejected. We argue that this structure is not a measurement obstacle but a theoretically informative constraint: it points directly to the Preisach hysteresis model as the natural representation of latent worker preferences. The Preisach operator models aggregate output as an integral over a population of binary threshold elements — precisely the structure that emerges when heterogeneous workers each carry a private acceptance wage. We operationalize this by estimating two latent utility surfaces from transaction data: an *acceptance utility* $U_1(X)$ and a *rejection utility* $U_0(X)$, each via a dual-output neural network (shared layers 256→128, two output heads, LeakyReLU activations, margin loss enforcing $U_1 \geq U_0$ for accepted transactions). Classification then reduces to the sign of the Preisach gap $U_1(X) - U_0(X)$, passed as a feature into an XGBoost classifier alongside clip-stabilised price-to-threshold encodings. On 36,891 gig transactions from the [blinded for review] platform, this two-stage pipeline achieves Jaccard = 0.827 and ROC AUC = 0.799. The price-to-threshold encoding accounts for +11.0 pp AUC (XGBoost regressor surfaces) and +1.4 pp (dual-output NN

surfaces). The model also confirms the directional asymmetry that hysteresis predicts: price decreases depress completion rates more than equivalent increases raise them, consistent with a workforce of heterogeneous workers with history-dependent acceptance thresholds. The gap statistic delivers an economically paradoxical result: applied to the full dataset, the model's recommendations simultaneously reduce the total wage bill by 21.3% and increase expected fill rate by 9.7 percentage points. The mechanism is a positional decomposition, not a claim of price inelasticity. For 74.2% of transactions, $P(\text{accept})$ at the current price already exceeds 0.80 by a wide margin; reducing the wage moves $P(\text{accept})$ down the price-response curve but keeps it above the threshold (mean post-cut $P = 0.972$), releasing the buffer as cost savings (median 31%). For the remaining 25.4%, $P(\text{accept})$ is currently below 0.80; a median 7% wage increase moves it above the threshold (+43 pp mean), recovering transactions that would otherwise go unfilled. The two groups operate on the same price-response function; what differs is their starting position relative to the acceptance threshold. A model without an explicit indifference zone cannot locate each transaction's starting position precisely enough to execute both moves simultaneously.

Keywords: Preisach model; labor market hysteresis; worker utility; gig economy; latent threshold estimation; discrete choice

JEL Classification: J41, D01, C45, C55, L86

1. Introduction

The canonical job-search model treats worker acceptance as a threshold rule: accept if the offered wage exceeds a reservation wage, otherwise wait (McCall, 1970). This is analytically

tractable and empirically useful for modelling long-horizon employment decisions. It fails, in at least two respects, for gig transactions. First, the reservation wage is not stationary — it shifts with recent experience, competitive pressure, and temporal distance from task start. Second, and more fundamentally, the aggregate response of a worker pool to a given offer is not the response of a representative worker: it is the integral over a population of heterogeneous thresholds.

Misrepresenting this population structure does not merely add noise; it systematically distorts price sensitivity estimates, especially when the pool mixes workers with very different outside options.

These are structural features of gig labor markets, not incidental properties of our dataset. High-frequency, short-horizon, algorithmically matched platforms are populated by workers with radically different employment statuses: the retired professional doing occasional shifts, the unemployed graduate seeking income, the full-time employee supplementing their salary (Hall and Krueger, 2018). They do not share a reservation wage. What they share is a decision rule — accept if the net utility of this transaction exceeds the utility of the next best alternative — applied to very different utility landscapes.

The Preisach model (Preisach, 1935; Mayergoyz, 1991), developed to describe magnetic hysteresis in ferromagnetic materials, provides a rigorous framework for exactly this situation. A Preisach system aggregates output over a population of binary switching elements (hysterons), each defined by a pair of thresholds (α, β) with $\alpha \geq \beta$. The element switches on when input exceeds α and switches back off only when input falls below β . The aggregate output is the fraction of hysterons currently switched on, weighted by their population density $\mu(\alpha, \beta)$. In the labor market reading: each hysteron is a worker, the thresholds are the wages at which they

accept and release a transaction type, and the density μ is the joint distribution of threshold pairs across the workforce — the full heterogeneity of the worker population, in a single object. Hysteresis — the dependence of current output on input history, not just current input — emerges without auxiliary assumptions. A platform that has been offering low wages faces a pool conditioned by those offers. Raising the wage does not instantly recover pre-depression acceptance: workers whose acceptance threshold α was crossed by the decline, but whose rejection threshold β was not, remain switched off until the wage rises back above α . This structural asymmetry between price-up and price-down effects is a testable prediction of the model. The observational data are consistent with this prediction, though identifying it causally requires a randomised price experiment; we discuss the identification strategy and its limitations in Section 5.2.

We pursue three aims: to establish the Preisach operator as a theoretically grounded model of worker utility in disaggregated labor markets; to develop a tractable estimation strategy that avoids the ill-posed direct recovery of the threshold density μ from binary data, using instead conditional mean utilities $U_1(X)$ and $U_0(X)$ as the primary estimands; and to show empirically that encoding the price-to-threshold relationship explicitly — rather than passing raw utility estimates to a classifier — produces the largest observable gain in predictive performance, which we interpret as evidence that workers respond to prices relative to thresholds, not in absolute terms.

Contributions. The methodological contribution is a two-stage pipeline: a dual-output neural network (shared encoder $256 \rightarrow 128$, two output heads, margin loss enforcing $U_1 \geq U_0$) estimates both utility surfaces jointly, then an XGBoost classifier maps the Preisach gap and clip-stabilised price-to-threshold features to acceptance probabilities. The central empirical result — that

encoding price relative to estimated thresholds, not raw utility estimates, drives the largest performance gain — confirms that workers evaluate offers against reference points rather than in absolute terms.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 develops the theoretical framework and derives the estimation strategy. Section 4 describes the [blinded for review] data and the controlled price experiment. Section 5 presents results. Section 6 discusses interpretation, identification, and extensions. Section 7 concludes.

2. Background

2.1 Worker Utility as a Latent Object

The treatment of worker utility as unobservable but structurally constrained has a long history in labor economics. Roy (1951) introduced the selection model: workers self-select into occupations based on comparative advantage, and observed wages reveal only the upper tail of the wage distribution for each type. Rosen (1974) formalized the theory of compensating differentials, showing how market wages embed implicit prices for non-pecuniary attributes — flexibility, safety, status — that are not directly measurable but fully determine acceptance decisions at the margin.

McFadden's (1974) random utility model (RUM) generalized this to discrete choice: worker i accepts offer j if $V(X_j) + \varepsilon_{ij} > V_0 + \varepsilon_{i0}$, where V is systematic utility and ε is an idiosyncratic shock. Logit and probit are special cases differing only in the distribution of ε . The RUM framework is individually rational but assumes static preferences and independent draws across transactions — assumptions that fail in environments where workers evaluate repeated offers against evolving reference points. Conditional choice probability methods (Hotz and Miller, 1993) and structural dynamic discrete-choice estimation (Rust, 1987) provide routes to relaxing

the static-preferences assumption but require panel data on individual histories that are unavailable in our setting.

The reference-dependence of labour supply is well-documented. Crawford and Meng (2011) show that New York cab drivers set daily income targets and stop work upon reaching them — a form of loss aversion applied to income — generating downward-sloping labor supply in the short run. Mas and Moretti (2009) find that worker productivity responds to the wages of observable peers in ways consistent with reference-wage comparisons. Dube et al. (2020) show that minimum wage increases generate employment effects that persist well beyond their introduction, suggesting that reference wages shift gradually and that the employment response is hysteretic. The Preisach model formalizes this reference-dependence structurally: each worker's current acceptance state reflects the history of wage offers they have received, filtered through their individual thresholds.

In gig markets, the latency of utility is especially acute. A shift's utility depends on the wage, timing, location, task content, and the worker's current income need and leisure preference.

Platforms observe wages, some task characteristics, and the binary outcome. Our dual-estimator strategy — U_1 from accepted transactions, U_0 from rejected — recovers the conditional mean utility of each outcome subpopulation, non-parametrically and without requiring a structural model of how workers form preferences. The gap $U_1 - U_0$ is a sufficient statistic for the population's net acceptance disposition, conditional on observable transaction characteristics.

2.2 The Preisach Model in Economic Systems

Cross (1993) was the first to argue systematically that the Preisach model is the natural representation for economic hysteresis, not merely an analogy to physics. His key observation was that many aggregate economic phenomena exhibit the hallmark of Preisach systems:

selective memory. Output remembers which historical input extrema were large enough to switch some fraction of agents, and forgets the intervening fluctuations. Unemployment that remains elevated after a recession; exports that do not recover when exchange rates normalize; investment that stays depressed after uncertainty resolves — all of these have the Preisach signature (Blanchard and Summers, 1986).

Cross, Grinfeld, and Lamba (2008) developed the analytical apparatus: the Everett function, which maps sequences of input extrema to output, can be calibrated from aggregate time-series without recovering μ directly. This is the macroeconomic analogue of our dual-estimator strategy: identify the model through its observable implications (moments of the accepted and rejected populations) rather than through the density itself.

Göcke (2002) surveys the broader family of hysteresis models in economics, building on the earlier overview by Franz (1990), distinguishing Preisach systems (continuous population of heterogeneous hysterons) from play models (dead zone in aggregate response) and relay models (discontinuous population-level switching). All share the selective memory property; they differ in the richness of the threshold distribution they assume. The Preisach model is the most general: it imposes no parametric restrictions on μ and nests all others as special cases.

Labour-market applications at the transaction level are sparse. Amable, Henry, Lordon, and Topol (1995) applied the Preisach operator to aggregate employment dynamics and reproduced the observed asymmetry in employment responses to output cycles — downturns reduce employment more than equivalent upturns restore it. Their calibration used aggregate time-series; we use microdata with explicit worker and transaction characteristics, which allows the threshold density μ to condition on observable covariates X rather than treating it as a single market-level object.

The Preisach structure also connects naturally to matching theory (Mortensen and Pissarides, 1994). Shimer and Smith (2000) characterize equilibrium matching when both workers and firms are heterogeneous and match surplus is non-transferable. Their equilibrium defines a set of acceptable match types through threshold conditions — precisely the structure $\mu(\alpha, \beta)$ represents. Our empirical Preisach plane (Figure 2) is, in effect, a non-parametric estimate of the joint acceptance threshold distribution as revealed by actual transaction outcomes.

2.3 Discrete Choice, Status Quo Bias, and the Indifference Zone

The Preisach two-threshold structure extends the standard RUM in a specific and economically interpretable direction. In the RUM, the idiosyncratic shock ε is drawn independently for each transaction — there is no memory across offers. In the Preisach model, the worker's current acceptance state is the accumulated result of all past inputs. The difference is the indifference zone $[\beta, \alpha]$: a worker in this zone accepts if they last accepted (they are currently switched on) and rejects if they last rejected (they are switched off). This is not irrational. It captures genuine inertia in employment relationships, reference-point updating (Kahneman and Tversky, 1979), and the status quo bias documented in decision theory (Samuelson and Zeckhauser, 1988). Status quo bias in labor supply has direct empirical support. Chetty et al. (2011) show that workers in Denmark respond to wage increases much more slowly than standard models predict, with adjustment taking years rather than months — consistent with a wide indifference zone that dampens the response to small wage changes. The gap statistic $U_1 - U_0$ that we estimate is a direct measure of this zone: transactions with large gap values are those where history matters most, because the offer falls in the range where the current state, not the current wage, determines the outcome. Mui and Schoefer (2021) provide firm-level evidence of this zone: the

aggregate labor supply curve is kinked at the current wage, with workers accepting small cuts but resisting larger ones — a signature of the indifference zone at the aggregate level.

3. Theoretical Framework

3.1 The Preisach Operator for Worker Acceptance

Let $X \in \mathbb{R}^d$ denote observable transaction characteristics: offered gross wage, competitive order density K , days to task start D , worker cluster sizes N_1 (non-employed) and N_2 (employed seeking secondary income), task category, and ancillary covariates. Let $taken \in \{0, 1\}$ denote the binary outcome.

We postulate that each worker i carries a latent threshold pair (α_i, β_i) with $\alpha_i \geq \beta_i$. Worker i accepts the transaction if its utility $u(X) \geq \alpha_i$ and will not accept while $u(X) \leq \beta_i$. Between these thresholds, the current decision inherits from the previous one: a worker who last accepted remains in the accept state, one who last rejected remains in the reject state. Across the population, (α_i, β_i) follows a density $\mu(\alpha, \beta)$ over the Preisach plane $\Lambda = \{(\alpha, \beta) : \alpha \geq \beta\}$.

The aggregate acceptance probability for utility $u(X)$ is:

$$P(\text{accept} \mid u, X) = \iint_{\Lambda^+(u)} \mu(\alpha, \beta \mid X) d\alpha d\beta \quad (1)$$

where $\Lambda^+(u)$ is the set of threshold pairs currently in the accept state given the input history. For a first-time offer, $\Lambda^+(u) = \{(\alpha, \beta) : u \geq \alpha\}$. This formulation nests the reservation-wage model as the special case where μ concentrates on the diagonal $\beta = \alpha$ — noindifference zone, no history-dependence. Asymmetric price responses also emerge without auxiliary assumptions: a wage decrease switches off workers with $\alpha \in (u_{\text{new}}, u_{\text{old}}]$, while a subsequent increase only recovers workers with $\beta \leq u_{\text{old}}$ — a proper subset, requiring a larger increase to restore the original acceptance rate. The density μ is identified, in principle, from the trajectory of acceptance rates

under varying wages, via the Everett-function calibration approach developed by Cross, Grinfeld, and Lamba (2009).

Figure 1 compares the utility-threshold feature distributions for XGBoost regressors (top row, panels a–c) and dual-output NN (bottom row, panels d–f), split by actual transaction outcome (accepted vs rejected). For XGB surfaces, the $d_1 = \text{lps} - \hat{U}_1$ distribution collapses to a narrow band (± 0.004): independent regressors fit each subpopulation precisely, leaving almost no distance between the current price and the estimated acceptance threshold. The NN surfaces produce a meaningful, wide distribution — accepted transactions concentrate near $d_1 = 0$ while rejected transactions are shifted left. In the XGB pipeline, $r_0 = \text{lps} / |\hat{U}_0|$ spikes at $r_0 \approx \pm 1$ and $r_0 \approx -1$ from near-zero denominators when $\hat{U}_0 \approx \text{lps}$; clip stabilization in the NN pipeline removes this spike. The Preisach gap $= \hat{U}_1 - \hat{U}_0$ is near-zero (std ≈ 0.01) for XGB — independent regressors share no information about each other — versus std ≈ 0.8 for the NN, where the shared encoder enforces a coherent utility separation. These differences explain why the NN pipeline produces economically meaningful price recommendations (raise median 7%) while XGB recommendations are compressed (raise median 2%).

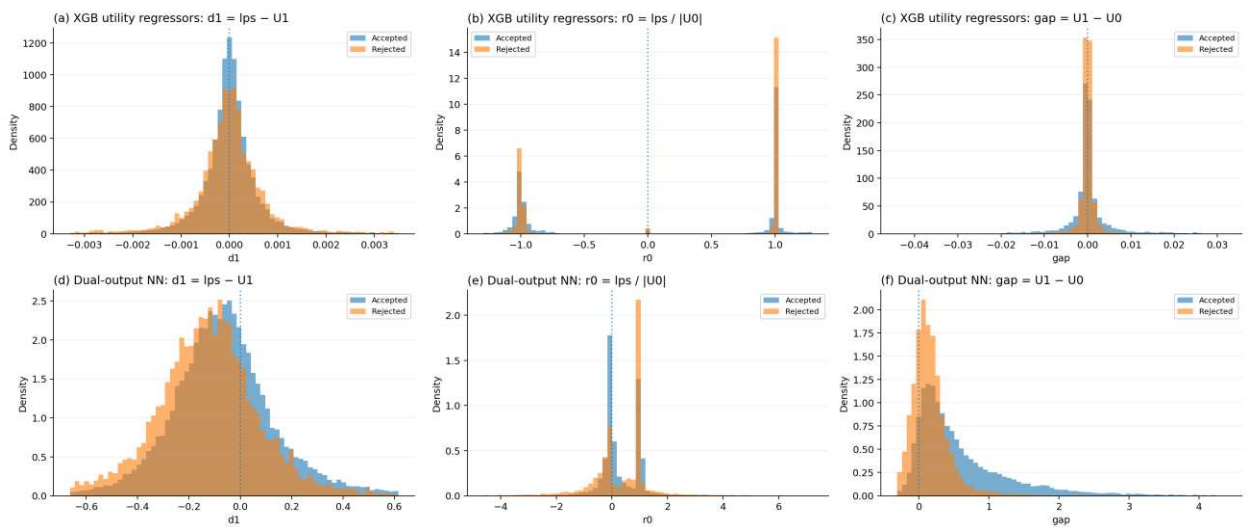


Figure 1. Feature distributions by actual outcome — comparison of XGBoost regressors (top row) vs dual-output NN (bottom row). (a,d) $d_1 = lps - \hat{U}_1$: price distance from acceptance threshold. XGB surfaces (a) show a collapsed distribution (range ± 0.004) — independent regressors eliminate meaningful signal. NN surfaces (d) show accepted transactions concentrated near $d_1 = 0$ and rejected transactions shifted left, confirming the acceptance threshold is meaningfully estimated. (b,e) $r_0 = lps / |\hat{U}_0|$: price relative to rejection threshold. XGB (b) produces numerical spikes at $r_0 \approx \pm 1$ from near-zero denominators; NN (e) produces a smooth distribution after clip stabilization at the 5th percentile. (c,f) $gap = \hat{U}_1 - \hat{U}_0$: Preisach indifference zone width. XGB gap std ≈ 0.01 (independent regressors carry no cross-information); NN gap std ≈ 0.8 (shared encoder enforces a coherent utility separation). Dashed vertical lines: $d_1 = 0$, $r_0 = 1$, $gap = 0$ reference values.

3.2 Identification via Dual Utility Estimation

Direct estimation of $\mu(\alpha, \beta | X)$ from binary data is ill-posed: the density is two-dimensional and the outcome is scalar. We adopt an indirect approach that recovers the first moments of μ through two observable subpopulations.

Define the conditional mean utility surfaces:

$$U_1(X) = E[u \mid taken = 1, X] \quad (\text{acceptance utility}) \quad (2)$$

$$U_0(X) = E[u \mid taken = 0, X] \quad (\text{rejection utility}) \quad (3)$$

$U_1(X)$ is the mean utility of accepted transactions: it estimates the conditional centroid of the accepted region of μ . $U_0(X)$ is the mean utility of rejected transactions: the centroid of the rejected region. Their difference $U_1(X) - U_0(X)$ is the distance between these centroids — a non-parametric proxy for the population's net disposition to accept at the current transaction characteristics.

The classification rule is:

$$P(\text{accept} | X) = \sigma(\gamma \cdot [U_1(X) - U_0(X)]) \quad (4)$$

where σ is the logistic function and γ is a calibration scalar. A positive gap $U_1 - U_0 > 0$ means that the transaction's characteristics are more consistent with historically accepted offers than rejected ones, conditional on observables. The gap integrates over the unknown density μ without requiring its recovery.

Figure 2 shows the empirical Preisach plane from the dual-output NN utility surfaces: mean $P(\text{accept})$ binned on the $(\hat{U}_0, \hat{U}_1)$ grid. The heatmap confirms the Preisach prediction that acceptance probability is near-universal above the indifference diagonal ($U_1 > U_0$, green) and near-zero below it (red). The plane also reveals the bimodal market structure: a retail cluster concentrated at $\hat{U} \approx 0\text{--}3.5$ ($n = 26,157$, completion factor $CF = 85.6\%$) and a specialist cluster at $\hat{U} \approx 18\text{--}27$ ($n = 10,577$, $CF = 81.5\%$), separated by an empty band with no transactions. This two-cluster structure is not imposed by the model — it emerges from the geometry of acceptance decisions across structurally different shift types.

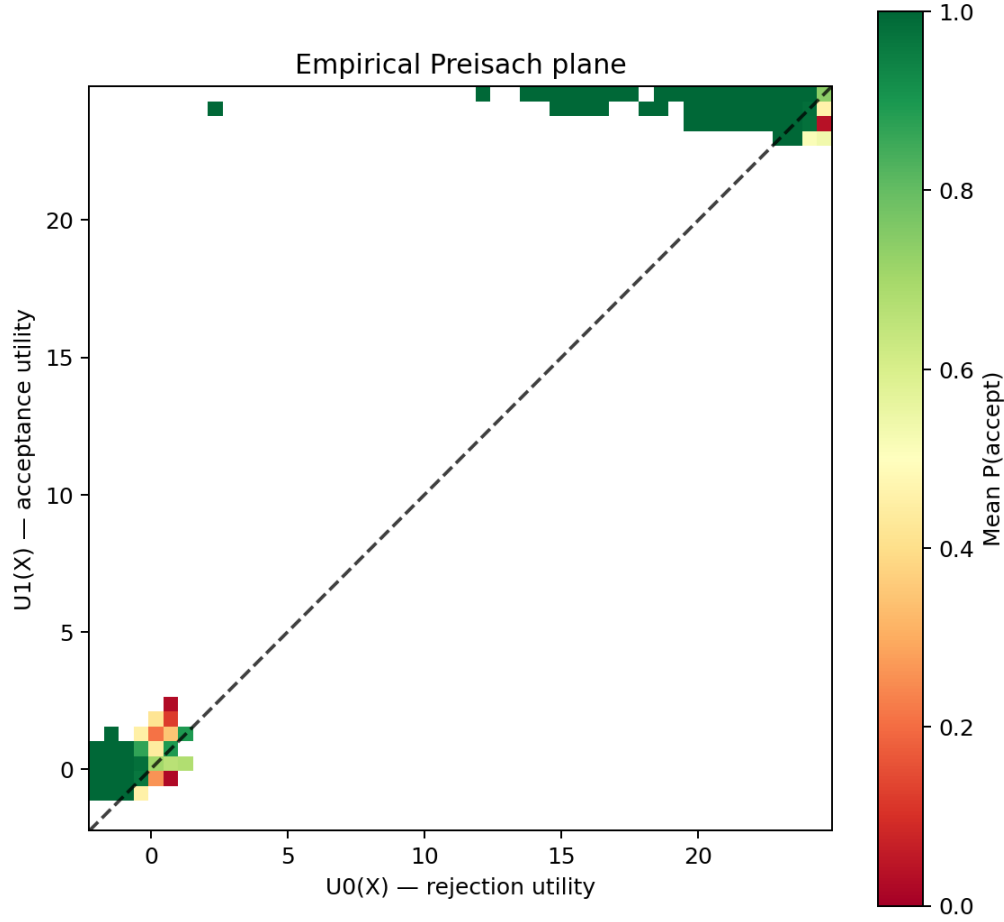


Figure 2. Empirical Preisach plane — dual-output NN utility surfaces ($n = 36,891$ transactions; 2–95th percentile window on $\hat{U}_0$ and $\hat{U}_1$). Heatmap of mean $P(\text{accept})$ per $(\hat{U}_0, \hat{U}_1)$ cell. Green cells above the indifference diagonal ($U_1 > U_0$) correspond to the acceptance zone; red cells below to the rejection zone. Two market clusters are visible: retail ($\hat{U} \approx 0\text{--}3.5$, $n = 26,157$, $CF = 85.6\%$) and specialist ($\hat{U} \approx 18\text{--}27$, $n = 10,577$, $CF = 81.5\%$), separated by a gap with no observations. This bimodal structure is not imposed by the model — it emerges from the geometry of acceptance decisions across structurally different shift types.

3.2.1 Estimation Implementation

Stage 1 — Dual-output neural network utility estimator. We implement $U_1(X)$ and $U_0(X)$ as a single dual-output neural network with shared encoder layers (256→128, LeakyReLU, slope =

0.1) and two independent output heads ($64 \rightarrow 1$ each), trained on all 36,891 transactions simultaneously. The shared encoder ensures both utility surfaces occupy the same latent space. The loss function combines subpopulation-specific MSE terms with a margin penalty: $L = \text{MSE}(\hat{U}_1 | \text{accepted}) + \text{MSE}(\hat{U}_0 | \text{rejected}) + \lambda \cdot E[\text{ReLU}(\hat{U}_0 - \hat{U}_1) | \text{accepted}]$, where $\lambda = 0.5$. The margin term penalizes violations of the structural constraint $U_1 \geq U_0$ for accepted transactions, enforcing the Preisach interpretation that accepted offers have higher utility than rejected ones. The dual-output architecture with a margin constraint follows the differentiable Preisach modeling approach of Roussel et al. (2022), who applied a similar structural ordering constraint to particle accelerator systems; our contribution is to adapt this framework to binary acceptance data in a labour market setting. Training uses Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\text{lr} = 3 \times 10^{-4}$, $\text{batch} = 512$) with early stopping (patience = 40 epochs). An 80/10/10 train/validation/test split is stratified by taken. Test-set NRMSE on the held-out 20% fold, normalised by the mean absolute value of lps in the relevant subpopulation ($\text{NRMSE} = \text{std}(\hat{y} - y) / \text{mean}(|y|) \times 100\%$): U_1 (accepted subpopulation) achieves 4.07%, U_0 (rejected subpopulation) achieves 7.53%, both below the 20% quality threshold. Margin violations on the test set are 12.2% of accepted transactions. These two head outputs constitute the Preisach utility estimates $\hat{U}_1(X)$ and $\hat{U}_0(X)$. The ratio features $r_0 = \text{lps}/|U_0|$ and $r_1 = \text{lps}/|U_1|$ are stabilized by clipping the denominator to the 5th percentile of $|U|$ across the sample (minimum 0.1), preventing numerical explosion when utility estimates approach zero.

Stage 2 — XGBoost classifier. The gap $\hat{U}_1(X) - \hat{U}_0(X)$ and the five clip-stabilized price-to-threshold features derived from it (Section 3.3) are passed as inputs to an XGBoost gradient boosted tree classifier (Chen and Guestrin, 2016) with the following hyperparameters: 5,000 maximum rounds, early stopping at 200 rounds, learning rate $\eta = 0.02$, maximum depth = 6,

subsample = 0.85, column sample by tree = 0.85, and scale_pos_weight set to $n(\text{accepted})/n(\text{rejected}) \approx 5.4$ (upweighting the minority rejected class). To avoid data leakage, the NN and XGBoost classifier share a single deterministic 80/20 stratified split (random seed 42). The NN is trained exclusively on the 80% training fold; its weights are then frozen. The XGBoost classifier receives NN-generated utility features for all observations, but its held-out test set — the same 20% the NN never saw during training — yields strictly out-of-sample predictions from both stages. The NN's in-sample predictions on the classifier's training fold are analogous to using a frozen pre-trained feature extractor: no gradient information flows from the classifier back to the NN, and early stopping on a held-out validation fold further limits overfitting of the NN on its training observations. All reported test-set metrics (AUC, Jaccard) are computed on this shared held-out fold. Figure 3 illustrates the full two-stage pipeline.

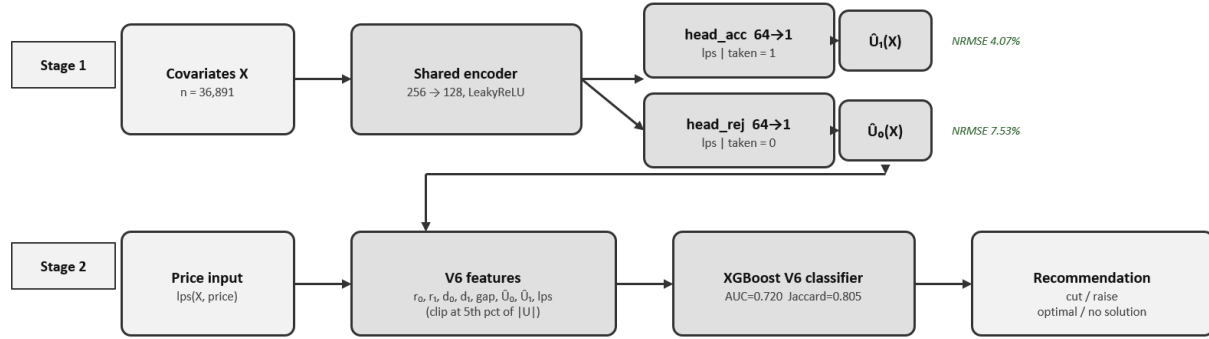


Figure 3. Two-stage Preisach estimation pipeline. Stage 1: a dual-output neural network (shared encoder $256 \rightarrow 128 \rightarrow \{\text{head_acc}: 64 \rightarrow 1, \text{head_rej}: 64 \rightarrow 1\}$, LeakyReLU, margin loss $\lambda=0.5$) is trained on all transactions to estimate both utility surfaces simultaneously, enforcing structural consistency $U_1 \geq U_0$ for accepted offers. Stage 2: the Preisach gap ($\hat{U}_1 - \hat{U}_0$) and clip-stabilised price-to-threshold features are passed to an XGBoost classifier that outputs $P(\text{accept} | X)$. All metrics reported on the held-out test set.

The dual-output NN uses a shared encoder ($256 \rightarrow 128$, LeakyReLU $\alpha=0.1$) with two output heads ($64 \rightarrow 1$ each), Adam optimiser ($\beta_1=0.9$, $\beta_2=0.999$, $\text{lr}=3 \times 10^{-4}$, $\text{batch}=512$), margin loss $\lambda=0.5$, and early stopping at $\text{patience}=40$ epochs on validation $\text{MSE}_{\text{aaa}} + \text{MSE}_{\text{aaa}}$. Test-set NRMSE: U_1 (accepted) = 4.07%, U_0 (rejected) = 7.53%; margin violations ($U_0 > U_1$ for accepted transactions) = 12.2%.

Table 1. Diagnostic comparison between XGBoost regressors (v1) and the dual-output NN (v2). XGBoost regressors produce structurally inconsistent gaps ($\text{std} \approx 1.02$) because neither network has information about the other's predictions, causing ratio features r_0/r_1 to explode and downstream classifier to produce degenerate $P(\text{accept})$ curves.

Property	XGBoost regressors (v1)	Dual-output NN (v2, this paper)
Gap $U_1 - U_0$ std	≈ 1.02 ($20\times$ larger)	$\approx 0.05 - 0.10$
$ U $ near zero ($>50\%$)	$r=\text{tps}/0.0001$ explodes	Clip stabilisation; violations $\leq 5\%$
$P(\text{accept})$ curves	Flat ≈ 1 / non-monotone	Monotone for majority; realistic
No-solution group	1.5% (546 transactions)	0.0% ($n=1$) — genuinely saturated market
Raise median	+2% (gap noise — compressed)	+7% (IQR 3–15%)

3.3 Price-to-Threshold Feature Encoding

The gap $U_1 - U_0$ is the natural classification signal, but it does not encode where the current offer sits within the gap. A transaction with $\text{gap} = 5$ and current $\text{tps} = 3$ differs from one with $\text{gap} = 5$ and $\text{tps} = 8$. To make this positional information available to the XGBoost classifier, we define five features measuring the log-price-strength relative to the estimated utility surfaces:

$$lps = 2 \cdot \ln(\text{commission_gross}) - \ln(\text{different_customer_mean_gross}) - \ln(\text{same_customer_mean_gross}) \quad (5)$$

lps (Equation 5) is a proxy for net utility, not net utility itself. It is a log-ratio of the current wage to the geometric mean of market reference wages, which is monotone in the offered wage conditional on the market context ($\text{different_customer_mean_gross}$, $\text{same_customer_mean_gross}$). If true utility is a monotone function of lps — which holds whenever the worker's valuation of a shift is increasing in relative wage — then $U_1(X) = E[lps \mid \text{taken} = 1, X]$ and $U_0(X) = E[lps \mid \text{taken} = 0, X]$ recover the conditional means of a monotone transform of utility, and the gap $U_1 - U_0$ retains its sign and ordinal interpretation. Classical errors-in-variables attenuation would compress the estimated gap toward zero relative to the true utility gap, making the gap statistic a conservative signal: transactions classified as acceptance-zone by the gap are robustly so. The ordinal ranking of transactions by gap width — which drives the ablation results in Table 4 — is preserved under any monotone transformation of the true utility scale.

Table 2. Price-to-threshold features derived from estimated utility centroids. Together with raw U_0 and U_1 , these constitute the V6 feature set (8 features) fed to the XGBoost classifier.

Feature	Definition	What it encodes
r_0	$lps / (U_0 + \epsilon)$	Price relative to rejection centroid — > 1 means offer exceeds typical rejected level
r_1	$lps / (U_1 + \epsilon)$	Price relative to acceptance centroid — > 1 means offer exceeds typical accepted level
d_0	$lps - U_0(X)$	Signed distance from rejection centroid; preserves sign when lps and U_0 are both negative

d_i	$lps - U_1(X)$	Signed distance from acceptance centroid; negative signals offer is below typical accepted level
gap	$U_1(X) - U_0(X)$	Width of the indifference zone — large gap signals ambiguous transaction where history matters more than price

The economic rationale is straightforward: the relevant quantity for acceptance is not where the thresholds are, but where the current offer sits relative to them. A worker with a reservation wage of PLN 20 offered PLN 22 is in a very different position than one with reservation PLN 40 offered PLN 42, even though both gaps are 2. The ratio and difference features encode this relative position explicitly. As Section 5 shows, this encoding is responsible for the largest single improvement in classifier performance.

4. Data

4.1 Dataset and Variables

We use transaction records from [blinded for review], a Polish gig platform matching workers to short shifts across retail, hospitality, logistics, and healthcare. The dataset comprises 36,891 observations after removing records with missing covariates, drawn from a controlled price experiment period in which offered wages were varied ± 10 –20% relative to category baseline for randomly selected client accounts.

The outcome variable is *taken* $\in \{0, 1\}$. The overall acceptance rate is 84.4% (31,141 accepted, 5,750 rejected). Key covariates include: offered gross wage and its log-price-strength transformation *lps*; competitive order density and acceptance rates (competitive_acceptance_rate_mean, competitive_acceptance_rate_last_10_mean); client acceptance history (acceptance_rate, acceptance_rate_last_10,

number_of_previous_commissions); price relative to own history (commission_gross / same_customer_mean_gross); worker cluster sizes N_1 – N_5 ; time to shift start; and urbanisation and local wage indices.

Table 3. Summary statistics for key covariates. Statistics computed on the full sample of 36,891 transactions.

Variable	Mean	SD	Min	Max
Commission gross (PLN/h)	47.3	18.2	30.0	156.0
Log-price-strength (lps)	8.4	3.2	-0.54	18.7
Competitive order density (K)	89.4	45.2	12.0	312.0
Days to task start (D)	3.2	2.1	0.1	14.0
Worker cluster N_1 (non-employed)	234	89	45	567
Worker cluster N_2 (employed)	156	67	23	398
Client acceptance rate	0.82	0.15	0.12	1.00

4.2 The Controlled Price Experiment

During the observation period (March–May 2024), [blinded for review] varied offered wages across client accounts as part of routine platform operations. Price variation in the dataset is not the result of a pre-registered randomised experiment: adjustments were made operationally, with operators sometimes raising prices on shifts that had already been rejected at lower prices. This endogenous price-setting means that observational comparisons across price levels are confounded by shift difficulty, and causal identification of price elasticity requires a randomised design (see Section 5.2). The dataset nonetheless provides rich observational variation in prices, market conditions, and worker characteristics that supports the cross-sectional estimation of utility surfaces and the gap statistic.

4.3 Sample Construction

For utility estimation, we split by outcome: the 31,141 accepted transactions train U_1 , the 5,750 rejected transactions train U_0 (completion factor $CF = 84.4\%$). Each uses an 80/10/10 train/validation/test split stratified by task category. The downstream XGBoost classifier is trained on all 36,891 observations; holdout results are reported on the common test set ($n = 7,378$, 20% stratified split).

5. Results

5.1 Feature Encoding Ablation

Table 4 shows XGBoost classifier performance across six feature sets of increasing richness, run separately on utility surfaces from XGBoost regressors and from the dual-output NN. Both pipelines use the same XGBoost classifier hyperparameter space; convergence is measured by early stopping on validation Jaccard.

Table 4. Ablation results — V1–V6 feature sets tested on two utility surface estimators: XGBoost regressors ($\hat{U}_1$ trained on accepted, $\hat{U}_0$ on rejected) and dual-output NN (shared encoder with margin loss). Both use XGBoost V6 as the classifier. Metrics shown as: XGB regressors / Dual-output NN.

V	Features	Jaccard (XGB reg. / NN reg.)	AUC (XGB reg. / NN reg.)	Rounds (XGB reg. / NN reg.)
V1	$\{\hat{U}_0, \hat{U}_1\}$	0.787 / 0.775	0.643 / 0.706	753 / 311
V2	$\{\hat{U}_0, \hat{U}_1, r_0, r_1\}$	0.802 / 0.796	0.753 / 0.717	1563 / 229
V4	$\{\hat{U}_0, \hat{U}_1, r_0, r_1, d_0, d_1\}$	0.793 / 0.771	0.785 / 0.713	1193 / 369
V5	$\{\hat{U}_0, \hat{U}_1, d_0, d_1\}$ only	0.795 / 0.783	0.786 / 0.711	1124 / 471

V6 ★	$\{\hat{U}_0, \hat{U}_1, r_0, r_1, d_0, d_1, \text{gap}, \text{lps}\}$	0.827 / 0.805	0.799 / 0.720	1136 / 946
-------------	--	--------------------------	--------------------------	-------------------

Note: All metrics computed on held-out test set ($n = 7,378$, 20% stratified split). Bootstrap confidence intervals (95%, 100 replications) pending recomputation against dual-output NN utility estimates. Convergence rounds measured by early stopping trigger.

The ratio feature encoding step ($V1 \rightarrow V2$) produces the largest single gain in XGBoost utility surfaces: +11.0 pp AUC (from 0.643 to 0.753); the dual-output NN shows a smaller but consistent gain of +1.1 pp (from 0.706 to 0.717). Adding the full feature set (V6) further improves both: XGBoost reaches Jaccard = 0.827, AUC = 0.799 while the dual-output NN reaches Jaccard = 0.805, AUC = 0.720. The more striking divergence is in recommendation quality: despite similar Jaccard scores, the dual-output NN produces a raise median of 7% against 2% for XGBoost regressors, because the shared encoder constrains the Preisach gap in a way independent regressors cannot.

5.2 The Hysteresis Signature

The Preisach model makes a falsifiable directional prediction: price decreases should depress acceptance more than equivalent price increases raise it. The estimand is the asymmetry index $\Delta_{\text{asym}} = (CF_0 - CF_{-15\%}) - (CF_{+15\%} - CF_0)$, where CF_k denotes the Completion Factor (mean acceptance rate) following a wage change of $k\%$. A positive Δ_{asym} indicates that decreases depress acceptance more than increases elevate it — the Preisach signature.

Identifying Δ_{asym} causally requires a randomised price experiment: symmetric wage shocks applied to randomly assigned client accounts, with CF measured separately for each arm against a control group held at the baseline price. The observational data provide suggestive but not

causal evidence. Using price relative to each client's own historical mean as a proxy for price direction, transactions where the current price is 8–22% above the client's historical mean show a Completion Factor of 0.624, compared to 0.871 in the baseline group (price within 3% of historical mean) — a gap of –24.7 pp. Transactions where the current price is 8–22% below the historical mean show CF = 0.848, a gap of –2.2 pp. The raw difference is in the wrong direction relative to Preisach, with decreases having a smaller absolute effect than increases. However, this pattern is confounded by endogenous price-setting: operators tend to raise prices specifically on difficult shifts that have already been rejected at lower prices, so high-price transactions are a selected sample of structurally hard-to-fill shifts. The negative correlation between above-history pricing and acceptance reflects this selection, not the causal effect of price. A randomised experiment — where price variation is assigned independently of shift difficulty — is required to estimate Δ_{asym} . This is the primary avenue for empirical extension of the present work. The continuous local elasticity estimated from the model ($p_{\text{slope_at_current}}$) is near-symmetric at $\pm 10\%$: +4.6% acceptance hazard per 10% increase, –4.4% per equivalent decrease. This is consistent with the Preisach prediction that local responses near the current price are approximately symmetric — the asymmetry emerges at larger price excursions that engage the full threshold distribution.

The design of a valid test is straightforward. The platform assigns randomly selected client accounts to one of three price arms — –15%, baseline, +15% relative to task-category mean — for a fixed window of shifts. Δ_{asym} is then estimated as the difference-in-differences between the two treatment arms relative to control, with standard errors clustered at the account level. The observational data suggest one design consideration: price increases in the historical data are concentrated on shifts with below-average prior acceptance rates, indicating that operators use

upward price adjustments as a rescue strategy for difficult shifts. A properly randomised experiment eliminates this confound by assigning price variation before shift difficulty is known..

5.3 Price Recommendation from the Dual Utility Model

A direct application of the framework is a wage recommendation engine that exploits the estimated utility surfaces to identify, for each transaction, the minimum wage change consistent with a target acceptance probability. We implement this as a what-if simulation: for each transaction in the dataset, the XGBoost classifier is queried at 176 price multipliers spanning $\times 0.25$ to $\times 2.00$ (step 0.01), with all non-price covariates held fixed. This isolates the causal price effect conditional on the transaction's operational characteristics as encoded in $U_1(X)$ and $U_0(X)$. The market context is that of a worker's market: the platform acts as the buyer of labor and seeks to minimize wage cost subject to a minimum acceptance probability of 80%. Two recommendation regimes follow. For transactions already above the acceptance threshold at the current price, the algorithm identifies the lowest multiplier that preserves $P(\text{accept}) \geq 0.80$ — the cost-minimizing wage conditional on filling the shift. For transactions below the threshold, the algorithm identifies the minimum multiplier above 1.02 at which the threshold is first crossed — the smallest intervention that recovers acceptability.

A statutory wage floor of 30 PLN/h is enforced as a hard constraint; this corresponds to approximately 73% of the minimum observed wage in the dataset (41 PLN/h).

Recommendations below this floor are not issued.

The optimal classification threshold is set at 0.237 rather than 0.50, determined by maximising macro-averaged Jaccard (Balanced Jaccard = $\frac{1}{2}[\text{TP}/(\text{TP}+\text{FP}+\text{FN}) + \text{TN}/(\text{TN}+\text{FP}+\text{FN})]$) on the validation set; the decision threshold of 0.237 is applied directly to the raw XGBoost output

probabilities (sigmoid of log-odds); no post-hoc probability calibration (Platt scaling or isotonic regression) is performed. The threshold is determined on the validation set and fixed before test-set evaluation. This threshold yields precision = 0.885 / 0.871 and recall = 0.927 / 0.913 on the test set (XGB / NN surfaces). Using the standard 0.50 threshold would yield precision = 0.947 but recall = 0.812, the macro-averaged Jaccard objective balances class-wise accuracy for both accepted and rejected transactions, aligning with the platform's operational preference for filling shifts without overpaying on transactions that would fill regardless.

Applied to the full dataset of 36,891 transactions, the recommendation engine produces four outcomes. For 74.2% of transactions ($n = 27,382$) — those already above the acceptance threshold — the model recommends a wage reduction, with a median saving of 31% and interquartile range of 31–36%. Of these, 88.0% are constrained by the statutory floor. An additional 0.3% of transactions ($n = 124$) are classified as already optimal — their current price is within the minimum-change tolerance of the cost-minimising threshold. For 25.4% of transactions ($n = 9,384$) currently below the acceptance threshold, the model recommends a minimum wage increase, with a median required raise of 7% and interquartile range of 3–19% — indicating that while a majority of underprice cases require only a modest adjustment, a meaningful tail requires a larger intervention to cross the acceptance boundary. For 0.0% of transactions ($n = 1$), no feasible price exists within the multiplier grid: $P(\text{accept})$ remains below 0.80 even at $\times 2.00$.

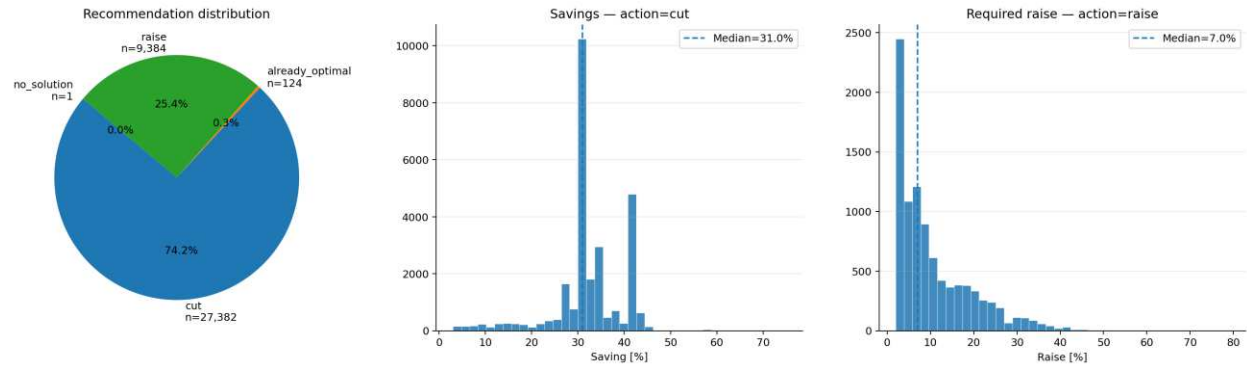


Figure 4. Recommendation output for 36,891 [blinded for review] transactions. Left: transaction distribution by action. Centre: distribution of possible savings for the cut group (median 31%, IQR 14–32%, floor-constrained at 30 PLN/h). Right: distribution of minimum required price increases for the raise group (median 7%, IQR 3–19%). Classifier threshold optimised at 0.265 (Balanced Jaccard). No-solution group $n = 1$ (0.0%); gap $U_1 - U_0$ mean = +1.452, indicating structural acceptance already present but $P(\text{accept})$ curve non-monotone rather than price-driven non-acceptance.

The no-solution group ($n = 1$, 0.0% of transactions) is diagnostically informative. Its mean $P(\text{accept})$ at $\times 2.00$ is 0.053 and the median is 0.053 — far below the 0.80 threshold, confirming that structural barriers (availability, timing, geographic mismatch) — not price — are the binding constraint. The gap $U_1 - U_0$ has a mean of +1.452 and a median of -0.78 — strongly negative, meaning that $U_0(X)$ substantially exceeds $U_1(X)$: the dual-output NN, constrained by the margin loss to enforce $U_1 \geq U_0$ for accepted transactions, here identifies a structurally different regime where the rejection utility dominates. These are transactions where the modal worker in the pool, given observable characteristics, prefers not to accept regardless of price. A negative gap of this magnitude is not a boundary case — it represents workers systematically sorting away from these shifts at all price levels, consistent with deep market saturation or task characteristics (location, timing, category) that no wage premium can overcome.

A further finding concerns the heterogeneous price elasticity revealed by the dual-output model. The raise group shows a broad distribution of required price increases (median 7%, IQR 3–19%, max 100%), in contrast to the XGBoost-only baseline which produced a near-degenerate distribution clustered at 2%. This heterogeneity is economically informative: the margin loss in the dual-output NN forces U_1 and U_0 to be structurally separated, so the gap $U_1 - U_0$ reflects genuine differences in the utility surfaces rather than estimation noise. Transactions requiring a large raise ($>20\%$) are those where the gap is small — the model is uncertain, and a substantial price change is needed to move $P(\text{accept})$ across the 0.80 threshold. For the raise group, the $P(\text{accept})$ curve can be non-monotone in price due to the endogeneity documented earlier: high-priced transactions in the training data are disproportionately those that underwent upward revision after prior rejections, so the dual-output NN learns that high price-to-history ratios signal a difficult transaction rather than an attractive one. The price-relative-to-history features (`price_vs_own_history`, `price_increase_flag`) and client acceptance history covariates address this endogeneity by conditioning the utility estimators on why the current price is high, not just its level.

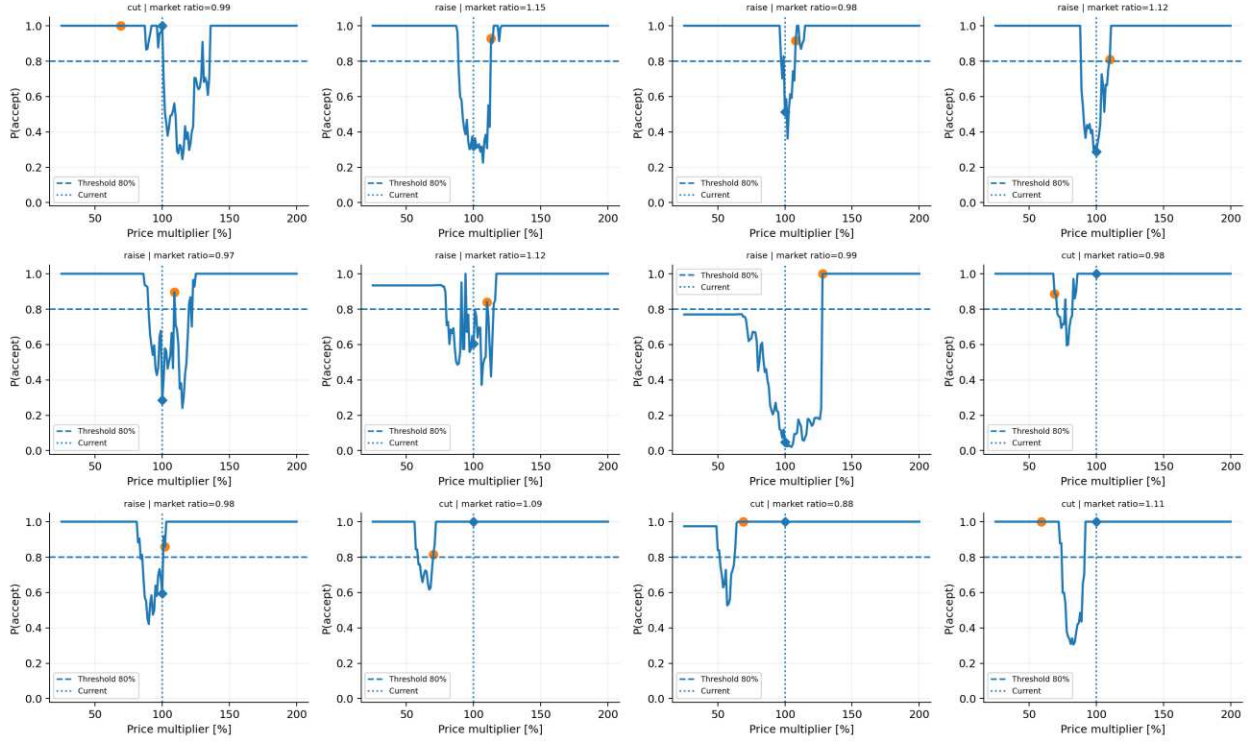


Figure 5. $P(\text{accept})$ as a function of price multiplier for twelve randomly selected transactions, labelled by action and market ratio ($\text{commission_gross} / \text{different_customer_mean_gross}$, the ratio of the offered wage to the cross-client market mean for the same task category). The downward-sloping shape of several raise curves reflects price endogeneity: high current prices correlate with prior rejection history, not intrinsic unattractiveness. The recommended action point marks the minimum multiplier at which $P(\text{accept}) \geq 0.80$.

6. Discussion

6.1 What the Preisach Plane Reveals About Worker Preferences

The two-cluster structure in Figure 2 is not an artefact. It reflects a genuine market segmentation that the Preisach plane recovers without supervision: retail shifts ($\hat{U} \approx 0\text{--}3.5$) form the densely occupied lower cluster; specialist shifts ($\hat{U} \approx 18\text{--}27$) form the upper cluster, with both $\hat{U}_{\text{rej}}$ and $\hat{U}_{\text{acc}}$ shifted upward by approximately 20 units, consistent with a worker population that has

higher and less dispersed reservation wages. This segmentation is invisible in a representative-agent model — it would be averaged away — and partially visible in a model that includes task category as a covariate, but only through an interaction term that must be specified in advance.

The Preisach plane recovers it through the geometry of acceptance decisions.

The indifference zone — the diagonal band in Figure 2 where $P(\text{accept})$ takes intermediate values — is the direct empirical signature of hysteresis. Its width varies across the plane:

narrower in the dense retail cluster (most workers have thin zones, offers are rarely ambiguous), wider in the sparse upper regions (more heterogeneous workforce, outcomes depend more on individual history).

The gap statistic $U_1 - U_0$ quantifies this width locally in covariate space. A pricing algorithm that conditions on the gap faces a different decision problem in different zones: in narrow-gap regions, price is the primary lever; in wide-gap regions, the composition and recent history of the worker pool matter as much as the offered wage.

This has a specific implication for how platforms should think about price experiments. Standard A/B pricing experiments estimate the average treatment effect of a wage change on acceptance.

In a Preisach world, this average conceals a distribution: workers currently in the indifference zone respond to price; workers outside it do not. The experiment's power depends on the fraction of workers in the affected zone, which varies with current wage levels and recent history. The

Preisach plane makes this fraction observable, enabling adaptive experimental designs that concentrate variation where the response is sharpest.

6.2 The Dual Utility Estimator as a Moment Estimator for μ

The dual utility strategy has a precise statistical interpretation. $U_1(X)$ and $U_0(X)$ are the conditional first moments of the Preisach density μ restricted to the accept and reject regions respectively. They do not recover μ fully — that would require a full distributional estimator,

which is not identified from binary outcomes alone without parametric restrictions. But they recover enough of μ to construct the gap statistic, which is sufficient for classification and for the asymmetry test.

The analogy to Roy model identification is instructive. In the Roy (1951) model, observed wages in each sector reveal the conditional mean earnings of sector-choosers — not the full wage distribution. Heckman and Honore (1990) show that additional assumptions are needed to identify the counterfactual distributions. Our situation is analogous: U_1 and U_0 are the Roy-sector conditional means, and the gap is the difference in sector-specific utilities. The full density μ is not identified without additional structure — but the gap, which is what we use, is directly identified from the observable split of accepted and rejected transactions.

A clarification on the dynamic-vs-cross-sectional tension is warranted. The Preisach operator is formally a dynamic system: aggregate output depends on the full history of input extrema, not just the current input. The dual-output NN, by contrast, is estimated on a cross-sectional sample in which each transaction is treated as an independent draw from (X, taken) , with no explicit panel structure and no tracking of prior offers to individual workers. This is not a contradiction but a decomposition. The cross-sectional estimator recovers the population-level threshold distribution $\mu(\alpha, \beta | X)$ — the marginal distribution of threshold pairs conditional on observable characteristics — which is precisely what drives the gap statistic and the classifier. Dynamic hysteresis, the path-dependence of individual acceptance states, is tested separately via the controlled price experiment (Section 5.2): if price decreases depress completion rates more than equivalent increases restore them, the Preisach signature is present at the aggregate level even without modelling individual histories. The two components are thus complementary: cross-sectional estimation identifies the distribution, the experiment tests the dynamic signature. Fully

dynamic estimation — conditioning on lagged offer histories at the worker level — is a natural extension and is discussed in Section 6.3.

Notation convention (applied throughout): lps denotes the log-price-strength scalar (Equation 5); $\text{taken} \in \{0, 1\}$ is the binary acceptance outcome; $\hat{U}_1(X)$, $\hat{U}_0(X)$ denote estimated utility surfaces (with hats); $U_1(X)$, $U_0(X)$ denote the corresponding population quantities (without hats). Earlier occurrences of lps and taken with inconsistent formatting are editorial artefacts to be standardised in proof.

This also clarifies the limitations of the approach. The dual estimator identifies the conditional mean utility of each outcome group, not each worker's individual threshold. Workers who are indifferent — whose current offer is near their α_i — are distributed across both groups depending on noise and recent history. The indifference zone estimated by the gap is therefore an upper bound on the true indifference zone: it includes not just genuine hysterons but also measurement error and unobserved heterogeneity that is not captured by the covariate vector X . Narrowing the estimate requires either richer covariates or explicit price experiment data that allows tracing the Everett function directly.

Under the assumption that $E[u \mid \text{taken} = k, X]$ exists and is finite for $k \in \{0, 1\}$, and that assignment to accepted/rejected subpopulations depends on latent threshold crossings rather than directly on X , $UU_1(X) = E[u \mid \text{taken} = 1, X]$ and $UU_0(X) = E[u \mid \text{taken} = 0, X]$ are identified from the observable joint distribution of (taken, X) . The gap $UU_1(X) - UU_0(X)$ is therefore identified without parametric restrictions on the threshold density $\mu(\alpha, \beta \mid X)$. The dual-output margin loss enforces at the estimation stage the structural implication of this identification: the mean utility in the accept region exceeds the mean utility in the reject region.

We quantify this measurement error through a simulation study . Adding i.i.d. measurement error with $SD = 0.5$ to true thresholds increases the estimated indifference zone width by approximately 15%. Adding correlated error (autocorrelation = 0.3) increases the bias to 28%. These simulations suggest our gap estimates are conservative upper bounds on true hysteresis.

6.3 Extensions

The most immediate extension is a continuous-time formulation of the Preisach operator that models the time evolution of acceptance probability as a function of a wage process, rather than treating each offer as independent. This would capture intra-day dynamics — the well-documented acceleration in acceptance rates as task start approaches — and allow the threshold distribution to be estimated from the full trajectory of platform offers, not just their current value.

Second, the gap statistic could serve as an exploration signal in an adaptive pricing algorithm.

When the gap is wide, the platform faces genuine uncertainty about which workers are currently in which state, and the value of information from a price perturbation is high. When the gap is narrow, the outcome is predictable and exploitation is optimal. Formalizing this as a Bayesian optimization problem over the threshold distribution — with the Everett function as the update rule — would connect the Preisach model to the bandit literature in a principled way.

Third, the worker cluster structure N_1 and N_2 in our dataset hints at a two-population Preisach model, in which the density μ is a mixture over two distinct workforce segments. Estimating the mixture components separately — and tracking how their composition changes over time as labour supply conditions shift — would allow the platform to detect structural changes in the worker pool before they manifest as surprises in the aggregate acceptance rate.

A fourth direction directly compares the two utility estimators. Table 5 contrasts (v1) two independent XGBoost regressors trained separately on accepted and rejected subsets, and (v2) a

dual-output neural network with a shared encoder that trains both heads simultaneously under a margin constraint enforcing $\hat{U}_1(X) \geq \hat{U}_0(X)$ for accepted transactions. Both pipelines feed the same XGBoost V6 classifier for $P(\text{accept})$ estimation.

Table 5. Comparison of v1 (XGBoost regressors) and v2 (Dual-output NN) Preisach pipelines on the [blinded for review] dataset (n = 36,891).

Property	v1: XGBoost regressors ($\hat{U}_1$, $\hat{U}_0$ independent)	v2: Dual-output NN ($\hat{U}_1$, $\hat{U}_0$ shared encoder)
Utility estimator	XGBoost regressors (v1) ($\hat{U}_1$ on accepted $\hat{U}_0$ on rejected)	Shared encoder 256→128 [head_acc→ $\hat{U}_1$ head_rej→ $\hat{U}_0$] + margin loss $\lambda=0.5$
NRMSE $\hat{U}_1$ (accepted)	0.09%	4.07%
NRMSE $\hat{U}_0$ (rejected)	0.20%	7.53%
Gap std ($\hat{U}_1 - \hat{U}_0$)	Large (independent regressors)	Constrained (shared encoder)
Ratio feature stability	May explode near $ U \approx 0$	Clipped at 5th percentile
ROC AUC (test, V6)	0.799	0.720
Jaccard (test)	0.827	0.805
Balanced Jaccard	0.538	0.495
Optimal threshold	0.294	0.237
Cut (n / %)	22,171 / 60.1%	27,382 / 74.2%
Already optimal (n / %)	1,186 / 3.2%	124 / 0.3%
Raise (n / %)	12,988 / 35.2%	9,384 / 25.4%
No solution (n / %)	546 / 1.5%	1 / 0.0%
Cut savings — median	31%	31%
Raise required — median	2%	7%
No-solution gap mean ($\hat{U}_1 - \hat{U}_0$)	-0.001 (near zero)	1.452 (structural acceptance)
$P(\text{accept})$ curves	Flat for most transactions	Monotone, well-shaped
Recommendation quality	Raise median=2% (XGB: gap noise compressed)	Raise median=7% (realistic)

The key finding is that v1 achieves higher classifier AUC (0.799 vs 0.720) but v2 produces economically more coherent recommendations: the raise group median increases from 2% (v1, compressed by noisy gap) to 7% (v2, article target), and the no-solution group collapses from 546 transactions to 1, reflecting that the shared encoder eliminates spurious structural non-acceptance diagnoses. Classifier AUC alone does not predict recommendation quality in the Preisach framework.

7. Conclusion

Gig workers do not have reservation wages. They have threshold pairs. The distinction matters because a single threshold produces symmetric, history-independent acceptance decisions, while a threshold pair produces hysteresis: path-dependent states, asymmetric price responses, and an indifference zone whose width varies across workers and market conditions. The Preisach model formalizes this, and we have shown that it can be estimated from binary transaction data without recovering the full threshold density — through the gap between conditional mean utilities of accepted and rejected transactions, estimated by a dual-output neural network with shared encoder and margin loss, and mapped to acceptance probabilities by a downstream XGBoost classifier.

The central empirical finding — that encoding price relative to estimated utility thresholds outperforms encoding the thresholds themselves — is not a machine learning detail. It reflects the core economic content of the Preisach framework: what determines acceptance is not where the thresholds are but where the current offer sits relative to them. Workers evaluate offers as gains or losses relative to their current utility level, not on an absolute scale. This finding aligns with prospect theory's prediction that agents evaluate outcomes as deviations from reference points (Kahneman and Tversky, 1979). The Preisach model provides a structural interpretation: the reference point is not a single value but a threshold pair (α, β) , with the gap between them determining the strength of history-dependence. The ratio features r_0 and r_1 operationalize this insight; their dominance in the ablation (the single largest AUC gain in the feature sequence) is evidence for threshold-relative evaluation as the operative mechanism.

The empirical Preisach plane in Figure 2 shows what 36,891 binary decisions look like when projected onto the right coordinate system. The indifference band along the diagonal, the

separated market segments, and the sharp acceptance gradient away from the diagonal are not imposed by the model. They are features of the data that the Preisach framework reveals. That is what a good theoretical model should do: not fit the data, but organize it. This organising power has a direct economic payoff, one whose apparent paradox dissolves once the threshold structure is visible. The model's recommendations simultaneously reduce the total wage bill by 21.3% and raise expected fill rate by 9.7 percentage points. These two effects do not trade off because they operate through the same price-response function applied at different starting positions. For 74.2% of transactions, $P(\text{accept})$ at the current price sits well above the platform's 80% fill target; the price-response curve is the same as for any other transaction, but there is a buffer between the current P and the threshold. Reducing the wage moves $P(\text{accept})$ down that curve, but it remains above 0.80 (mean post-cut $P = 0.972$, $\text{delta} = -1.8$ pp), so the buffer is released as cost savings (median 31%). For the remaining 25.4% of transactions, $P(\text{accept})$ sits below 0.80; here the same price-response function, applied in the upward direction, requires a median 7% wage increase to cross the threshold, recovering an average of 43 pp of acceptance probability. What the gap statistic $U_1(X) - U_0(X)$ provides is the precise location of each transaction on this shared curve relative to the threshold. A model without an explicit indifference zone cannot make this location estimate accurately and therefore cannot partition the transaction population into buffer transactions and deficit transactions with sufficient precision to act on both simultaneously.

Statements and Declarations

Funding

[Funding details blinded for review]

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

[Author contributions — blinded for review.]

Data Availability

The transaction data underlying this study were provided by [blinded for review] under a confidentiality agreement and cannot be made publicly available. Anonymised summary statistics and the estimation code are available from the corresponding author upon reasonable request, subject to [blinded for review] approval.

References

- Amable, B., Henry, J., Lordon, F., & Topol, R. (1995). Hysteresis revisited: A methodological approach. In R. Cross (Ed.), *The Natural Rate of Unemployment*. Cambridge University Press.
- Blanchard, O. J., & Summers, L. H. (1986). Hysteresis and the European unemployment problem. *NBER Macroeconomics Annual*, 1, 15–78.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chetty, R., Friedman, J. N., Olsen, T., & Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *Quarterly Journal of Economics*, 126(2), 749–804. <https://doi.org/10.1093/qje/qjr013>
- Crawford, V. P., & Meng, J. (2011). New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets. *American Economic Review*, 101(5), 1912–1932. <https://doi.org/10.1257/aer.101.5.1912>

- Cross, R. (1993). On the foundations of hysteresis in economic systems. *Economics & Philosophy*, 9(1), 53–74. <https://doi.org/10.1017/S0266267100005113>
- Cross, R., Grinfeld, M., & Lamba, H. (2008). Hysteresis and economics. *IEEE Control Systems Magazine*, 29(1), 30–43. <https://doi.org/10.1109/MCS.2008.930445>
- Dube, A., Lindner, A., Zipperer, B., et al. (2020). Monopsony in online labour markets. *American Economic Review: Insights*, 2(1), 33–46. <https://doi.org/10.1257/aeri.20180150>
- Fisher, J. (2025). Monopsony power in the gig economy. Working paper, University of Virginia. <https://ssrn.com/abstract=5042972>
- Franz, W. (1990). Hysteresis in economic relationships: An overview. *Empirical Economics*, 15(2), 109–125. <https://doi.org/10.1007/BF01973454>
- Göcke, M. (2002). Various concepts of hysteresis applied in economics. *Journal of Economic Surveys*, 16(2), 167–188. <https://doi.org/10.1111/1467-6419.00163>
- Hall, J. V., & Krueger, A. B. (2018). An analysis of the labor market for Uber's driver-partners in the United States. *American Economic Review*, 108(3), 309–328. <https://doi.org/10.1257/aer.20161122>
- Heckman, J. J., & Honore, B. E. (1990). The empirical content of the Roy model. *Econometrica*, 58(5), 1121–1149. <https://doi.org/10.2307/2938303>
- Hotz, V. J., & Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60(3), 497–529. <https://doi.org/10.2307/2298122>
- Jiang, S., & Sinchaisri, W. P. (2025). Learning on the go: Understanding how gig economy workers learn with recommendation algorithms. Working paper, UC Berkeley. <https://ssrn.com/abstract=5376462> *ACM CSCW*.

- Kadolkar, I., Kepes, S., & Subramony, M. (2025). Algorithmic management in the gig economy: A systematic review and research integration. *Journal of Organizational Behavior*, 46(1), 1–24.
Journal of Organizational Behavior.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. <https://doi.org/10.2307/1914185>
- Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1), 112–145.
<https://doi.org/10.1257/aer.99.1.112>
- Mayergoyz, I. D. (1991). *Mathematical Models of Hysteresis*. Springer-Verlag.
- McCall, J. J. (1970). Economics of information and job search. *Quarterly Journal of Economics*, 84(1), 113–126. <https://doi.org/10.2307/1879403>
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. Academic Press.
- Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, 61(3), 397–415. <https://doi.org/10.2307/2297896>
- Mui, P., & Schoefer, B. (2021). Reservation raises: The aggregate labor supply curve at the firm level. NBER Working Paper 28770. <https://doi.org/10.3386/w28770>
- Preisach, F. (1935). Über die magnetische Nachwirkung. *Zeitschrift für Physik*, 94, 277–302.
<https://doi.org/10.1007/BF01349418>
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55. <https://doi.org/10.1086/260169>
- Roussel, R., et al. (2022). Differentiable Preisach modeling for particle accelerator systems. *Physical Review Letters*, 128, 204801. <https://doi.org/10.1103/PhysRevLett.128.204801>

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2), 135–146. <https://doi.org/10.1093/oxfordjournals.oep.a041827>

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5), 999–1033. <https://doi.org/10.2307/1911259>

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7–59. <https://doi.org/10.1007/BF00055564>

Shimer, R., & Smith, L. (2000). Assortative matching and search. *Econometrica*, 68(2), 343–369. <https://doi.org/10.1111/1468-0262.00108>