© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

Behavioral Anomaly Detection in Social Security Claims Using a Longitudinal Statistical Profiling Approach with Income Variance Metrics

OMETAN S. OLOKOR[1], KAYOH O. CLINTON[2], EKUMA-OKEREKE EYINNAYA[3], OKEDOYE M. AKINDELE[4]

[1,2,3,4]*Federal University of Petroleum Resources, Effurun, Nigeria*

Abstract- ***Social security fraud imposes substantial fiscal and equity costs on pension systems worldwide, yet systematic statistical methods for its detection remain underdeveloped, particularly in resource limited institutional settings. This paper develops and validates a longitudinal behavioral profiling framework for detecting anomalous claims in a social security context. We introduce the Income Variance Ratio (IVR), a metric quantifying the proportional divergence between an individual's documented earnings trajectory and their claimed benefit entitlement, and demonstrate its theoretical and empirical superiority over conventional scalar risk scores. Building on a formal probabilistic model of claim generation, we derive a set of structural theorems characterizing the stochastic separation between legitimate and fraudulent claim populations, and establish consistency and convergence guarantees for the proposed estimators. Empirically, we apply the framework to a longitudinal dataset of 10,000 social security claims spanning 18 months, of which 746 (7.46%) were confirmed fraudulent. All eight primary features exhibit statistically significant distributional separation (Mann-Whitney $p < 10{-}13$; Kolmogorov-Smirnov $p < 10{-}10$). The IVR achieves the largest rank-bi-serial correlation ($r = 0.599$) among all features. A Random Forest classifier trained on the full feature set attains a cross-validated AUC of 0.826 and a perfect confusion matrix at a threshold of 0.40 on held-out data. Critically, IVR decile analysis reveals a non-linear, threshold-crossing pattern in which the top decile concentrates 42.2% fraud prevalence against a baseline of 7.46%, corresponding to a 5.7-fold lift. The Documentation Completeness × Geographic Risk interaction surface exhibits cells with fraud rates exceeding 100% in small-sample high-risk strata. These findings provide both theoretical grounding and practical guidelines for deploying income-variance-based fraud screening in pension and social security administration.***

Keywords: ***social security fraud detection; income variance ratio; behavioral anomaly scoring; longitudinal statistical profiling; random forest classification; Mann-Whitney test; principal component analysis; unsupervised anomaly detection.***

I. INTRODUCTION

Social security fraud is the submission of falsified, inflated, or otherwise illegitimate claims for pension or benefit disbursements represents a global governance challenge with documented annual losses running to billions of dollars across national pension systems (World Health Organization, 2019; Pickett, 2011; Albrecht et al., 2012). The problem is particularly acute in developing economies where contributory records are fragmented, identity verification infrastructure is nascent, and institutional investigative capacity is constrained (van der Berg et al, 1995). Despite growing interest in data-driven fraud detection within financial services (Phua et al., 2010; Bhattacharyya et al., 2011; Baesens et al., 2015), systematic statistical frameworks tailored to the administrative and actuarial peculiarities of social security systems remain comparatively scarce.

Existing approaches to social security fraud detection divide broadly into three families. Rule-based systems apply hard thresholds on individual feature: age, claim amount, contribution history and suffer from low recall as fraudsters adapt to known screening criteria (Bolton and Hand, 2002; Ngai et al., 2011). Supervised machine learning methods, including logistic regression, decision trees, and ensemble methods, have demonstrated strong discriminative performance in credit card fraud (Bhattacharyya et al., 2011; Dal Pozzolo et al., 2014; Carcillo et al., 2018) and healthcare billing fraud (Joudaki et al., 2015; Bauder and Khoshgoftaar,

2017), but their application to pension claims is limited by severe class imbalance and the absence of large labeled training corpora (Chawla et al., 2002; He and Garcia, 2009). Unsupervised anomaly detection methods, isolation forests (Liu et al., 2008), autoencoders (Chalapathy and Chawla, 2019), and local outlier factors (Breunig et al., 2000) circumvent the labeling requirement but lack interpretability and principled calibration for regulatory use (Chandola et al., 2009).

This paper addresses these gaps through these complementary contributions. We develop a formal probabilistic model of claim generation under both legitimate and fraudulent regimes. Within this model we define the Income Variance Ratio (IVR) as a sufficient statistic for income-trajectory anomaly, derive its asymptotic distribution, and prove a separation theorem establishing that IVR provides consistent discrimination between the two regimes as sample size grows. We further derive a Composite Anomaly Score (CAS) that aggregates multiple behavioral signals with provable monotonicity properties, and characterize the geometry of the fraud boundary in the feature-space principal components. We propose a five-stage longitudinal profiling pipeline: (i) feature engineering and IVR construction, (ii) non-parametric distributional testing, (iii) IVR decile stratification, (iv) supervised classification with threshold calibration, and (v) interaction risk surface estimation. Each stage is grounded in corresponding formal propositions that connect the statistical procedures to the underlying probabilistic model. We validate the framework on a longitudinal administrative dataset of 10,000 social security claims spanning 18 months. The dataset represents a multi-region, multi-bureau collection covering three geographically distinct administrative zones with independently verified fraud labels, providing an unusually clean evaluation environment. Our empirical results confirm the theoretical predictions, with IVR emerging as the single strongest discriminator and the Random Forest classifier achieving perfect hold-out classification at a threshold chosen by calibration analysis.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 presents the formal probabilistic model. Section 4 defines the derived anomaly metrics and their theoretical properties. Section 5 describes the empirical methodology. Section 6 characterizes the dataset. Section 7 presents all empirical results. Section 8 discusses implications. Section 9 concludes.

II. RELATED LITERATURE

The statistical detection of fraud has deep roots in actuarial science and forensic accounting. Benford (1938) noted that the leading digit distribution of naturally-occurring figures deviates from uniform, a property later formalized and applied to financial fraud detection by Nigrini (1999). Bolton and Hand (2002) provide a comprehensive survey of statistical methods for fraud detection, distinguishing supervised classification from unsupervised deviation scoring. The theoretical underpinnings of anomaly detection as hypothesis testing are reviewed in Chandola et al. (2009), who formalize the problem as testing a null of normality against a composite alternative. Our framework builds on this hypothesis-testing perspective while extending it to longitudinal behavioral profiles.

Bhattacharyya et al. (2011) demonstrate that Random Forest and Support Vector Machine classifiers substantially outperform logistic regression on highly imbalanced credit card fraud data. Dal Pozzolo et al. (2014) introduce the Self-Organizing Map combined with Multi-Layer Perceptron approach for streaming credit fraud, emphasizing the role of concept drift. Carcillo et al. (2018) address the operational challenge of reducing false alarms through SCARFF, a streaming resampling framework. Baesens et al. (2015) provide a comprehensive treatment of analytics for fraud management, covering model evaluation metrics appropriate for imbalanced problems. The Random Forest methodology we employ builds on Breiman (2001) and its extensions to class-imbalanced settings (Chen et al., 2004).

The closest methodological analogues to social security fraud detection arise in healthcare billing. Joudaki et al. (2015) review statistical and machine learning methods for health insurance fraud,

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

cataloguing feature engineering approaches that parallel our income-consistency and documentation completeness metrics. Bauder and Khoshgoftaar (2017) develop a claim-level profiling approach for Medicare fraud that combines physician-level peer comparison with individual behavioral trajectory analysis, a design philosophy closely aligned with our longitudinal framework.

A smaller literature specifically examines income misreporting in social benefit contexts. Pudney et al. (2004) models income underreporting in benefit claims as a mixture of truthful and strategic reports, deriving identification conditions for mixture parameters from the observed income distribution. Yaniv (1997) establishes game-theoretic conditions under which rational agents misreport income given audit probabilities, a framework that motivates our use of income variance as a behavioral signal. Our IVR metric operationalizes these insights in an administrative claims context.

Liu et al. (2008) introduce the Isolation Forest algorithm, exploiting the observation that anomalies require fewer random partitions to isolate than normal observations. The algorithm provides our unsupervised baseline in Section 7. Breunig et al. (2000) propose Local Outlier Factor (LOF) as a density based score; its theoretical connection to our CAS is established in Proposition 4.3 below. Chalapathy and Chawla (2019) survey deep learning approaches to anomaly detection, including autoencoders and variational autoencoders, whose representational power exceeds the linear dimensionality reduction in our PCA analysis but at substantially greater computational cost.

III. PROBABILISTIC MODEL OF CLAIMS

3.1 Feature Space and Notation

Let $\mathcal{X} \subset R^p$ denote the feature space for a social security claim, where p is the number of observable features. For each claim $i \in \{1, \dots, n\}$, we observe a feature vectorx_$x_i = (x_{i1}, \dots, x_{ip})^\top \in \mathcal{X}$ together with a latent fraud label Y_i∈{0,1}, where Y_i=1 denotes a fraudulent claim. Throughout the paper we consider the eight primary features: Claimant Age (A), Claim Amount (C), Years of Contributions (K), Income Consistency (ρ), Geographic Risk Score (G), Documentation Completeness (D), Claim Timing in Days (T), and Previous Claims Count (N).

Definition 3.1 (Claim Feature Vector). The claim feature vector for claim i is

$$\boldsymbol{x_i} = (A_i, C_i, K_i, \rho_i, G_i, D_i, T_i, N_i)^\top \in \mathcal{X} \subset R^8.$$

The legitimate feature distribution is $F_0 = \mathbb{P}(\boldsymbol{x} \mid Y = 0)$ and the fraudulent feature distribution $F_1 = \mathbb{P}(\boldsymbol{x} \mid Y = 1)$.

Assumption 3.2 (Mixture Model). The marginal distribution of x satisfies

$$F = (1 - \pi)F_0 + \pi F_1,$$

where $\pi = \mathbb{P}(Y = 1)$ is the fraud prevalence rate. We assume $0 < \pi < 1$ is strictly positive and fixed.

Assumption 3.3 (Stochastic Dominance). For each risk-increasing feature $j \in \mathcal{J}_+ = \{C, G, T, N\}$, the fraudulent marginal distribution first-order stochastically dominates the legitimate marginal: $F_{1,j}(t) \le F_{0,j}(t)$ for all $t \in \mathbb{R}$. For each risk-decreasing feature $j \in \mathcal{J}_- = \{A, K, \rho, D\}$, the reverse holds: $F_{0,j}(t) \le F_{1,j}(t)$.

Assumption 3.3 is consistent with the empirical patterns documented in Table 1: fraudulent claimants are younger $\overline{A}_1 < \overline{A}_0$, have fewer contribution years $\overline{K}_1 < \overline{K}_0$, lower income consistency $\overline{\rho}_1 < \overline{\rho}_0$, and lower documentation completeness, while claiming larger amounts, having higher geographic risk, later timing, and more prior claims.

3.2 The Income Trajectory Model

Let $\{W_{it}\}_{t=1}^{K_i}$ denote the sequence of documented annual earnings for claimant i over K_i contribution years. We model the true earnings trajectory as:
Definition 3.4 (Earnings Process). For a legitimate claimant, annual earnings follow an AR(1) process with trend:The Income Trajectory Model

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

$$W_{it} = \mu_i + \beta t + \varepsilon_{it}, \quad \varepsilon_{it} = \phi\varepsilon_{i,t-1} + \xi_{it}, \quad \xi_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2),$$

where $|\phi| < 1$ ensures stationarity. The long-run mean income is $\bar{W}_i = \mu_i + \beta(K_i + 1)/2$ and the long-run variance is $\sigma_w^2/(1-\phi^2)$.

Definition 3.5 (Income Variance Ratio). Let $\hat{\sigma}_{W,i}^2 = (K_i - 1)^{-1} \sum_{t=1}^{K_i} (W_{it} - \bar{W}_i)^2$ denote the sample variance of the earnings sequence and let C_i denote the claimed benefit amount. The Income Variance Ratio for claim i is

$$IVR_i = \frac{\hat{\sigma}_{W,i}^2}{C_i/K_i} = \frac{(K_i - 1)^{-1} \sum_{t=1}^{K_i} (W_{it} - \bar{W}_i)^2}{C_i/K_i}.$$

The denominator C_i/K_i is the implied average annual entitlement, i.e., the benefit amount normalised by contribution years. A large IVR thus indicates that income was highly volatile relative to the implied entitlement, signalling either strategic inflation of C_i or misrepresentation of K_i.

Lemma 3.6 (IVR under the Legitimate Regime). Under the legitimate earnings process (Definition 3.4), the IVR concentrates near a baseline value η_0 that depends only on model parameters:

$$IVR_i \xrightarrow{p} \eta_0 = \frac{\sigma_w^2/(1-\phi^2)}{\mu_i + \beta(K_i + 1)/2} \quad as \ K_i \to \infty.$$

Proof. By the law of large numbers, $\hat{\sigma}_{W,i}^2 \xrightarrow{p} \sigma_w^2/(1-\phi^2)$ as $K_i \to \infty$ for the stationary AR(1) process. Since $C_i/K_i \to \bar{W}_i = \mu_i + \beta(K_i + 1)/2$ under the legitimate pricing rule $C_i = K_i \bar{W}_i$, the result follows by the continuous mapping theorem.

Lemma 3.7 (IVR Inflation under the Fraudulent Regime). Suppose a fraudulent claimant reports a claim amount $C_i = (1 + \delta_i) K_i \bar{W}_i$ with inflation factor $\delta_i > 0$, while the variance of the underlying earnings sequence remains $\hat{\sigma}_{W,i}^2 \approx \sigma_w^2/(1-\phi^2)$. Then

$$IVR_i \approx \frac{\eta_0}{1 + \delta_i}.$$

However, if the fraudulent claimant also fabricates or truncates the contribution sequence, introducing variance $\hat{\sigma}_{W,i}^2 = (1 + \gamma_i)\sigma_w^2/(1-\phi^2)$ with $\gamma_i > \delta_i$, then $IVR_i > \eta_0$.

Proof. Substituting $C_i = (1 + \delta_i) K_i \bar{W}_i$ into Definition 3.5:

$$IVR_i = \frac{\hat{\sigma}_{W,i}^2}{C_i/K_i} = \frac{\sigma_w^2/(1-\phi^2)}{(1 + \delta_i)\bar{W}_i} = \frac{\eta_0}{1 + \delta_i}.$$

For the second case, $\hat{\sigma}_{W,i}^2 = (1 + \gamma_i)\sigma_w^2/(1 - \phi^2$, so $IVR_i = (1 + \gamma_i)\eta_0/(1 + \delta_i) > \eta_0$ iff $\gamma_i > \delta_i$

Remark 3.8. Lemma 3.7 implies that IVR inflation is driven by income fabrication (large γ_i) rather than mere claim inflation (large δ_i only). This is behaviorally significant: fabricated contribution histories introduce earnings volatility that is measurably inconsistent with legitimate career trajectories.

3.3 Separation Theorem

Theorem 3.9 (Stochastic Separation of IVR Distributions). Under Assumptions 3.2–3.3 and the earnings model of Definition 3.4, the IVR distributions under the two regimes satisfy:

$$F_{1,IVR}(t) \leq F_{0,IVR}(t) \quad forall \ t \geq 0,$$

with strict inequality on a set of positive measure. That is, IVR first-order stochastically dominates under the fraudulent regime.

Proof. From Lemma 3.7, the fraudulent IVR satisfies $IVR_i^{(1)} = (1 + \gamma_i)\eta_0/(1 + \delta_i)$. Since fraudulent claims involve systematic income fabrication $\gamma_i > 0$ and not merely benefit inflation, the distribution of $IVR^{(1)}$ places more mass on large values than $IVR^{(0)} \approx \eta_0$. Formally, for any $t \geq 0$:

$$\mathbb{P}(IVR^{(1)} > t) = \mathbb{P}\left(\frac{1+\gamma}{1+\delta} > \frac{t}{\eta_0}\right) \geq \mathbb{P}\left(\frac{1+\gamma}{1+\delta} > \frac{t}{\eta_0} \mid \gamma = \delta\right) = \mathbb{P}(IVR^{(0)} > t),$$

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

where the inequality is strict for $t > \eta_0$ because $\mathbb{P}(\gamma > \delta) > 0$ by assumption.

Corollary 3.10 (IVR as a Consistent Discriminator). Let $\hat{F}^n_{0,IVR}$ and $\hat{F}^n_{1,IVR}$ denote the empirical CDFs of IVR in the legitimate and fraudulent subsamples of size n_0 and n_1 respectively. The Mann-Whitney statistic $U_n = \mathbb{P}(IVR^{(1)} > IVR^{(0)})$ satisfies $U_n \xrightarrow{p} U^* > 1/2$ as $\min(n_0, n_1) \to \infty$, confirming that IVR is a consistent discriminator in the sense of Bamber (1975).

Proof. By Theorem 3.9, $F_{1,IVR} \neq F_{0,IVR}$, so $U^* = \mathbb{P}(IVR^{(1)} > IVR^{(0)}) \neq 1/2$. The stochastic dominance direction gives $U^* > 1/2$. Consistency of U_n follows from the Glivenko-Cantelli theorem applied to the empirical CDFs.

IV. DERIVED ANOMALY METRICS

4.1 Composite Anomaly Score (CAS)

Definition 4.1 (Composite Anomaly Score). Let $\tilde{x}_{ij}$ denote the $[0,1]-$ normalised value of feature j for claim i with direction chosen so that higher values indicate greater anomaly. The Composite Anomaly Score is the weighted mean:

$$CAS_i = \sum_{j=1}^{p} w_j \tilde{x}_{ij}, \quad \sum_{j=1}^{p} w_j = 1, \; w_j \geq 0.$$

Proposition 4.2 (Monotonicity of CAS). The CAS is monotone non-decreasing in each feature in the direction of fraud. That is, for any claim i and any feature j, increasing $\tilde{x}_{ij}$ (in the anomaly direction) does not decrease CAS_i.

Proof. Immediate from the linearity of the weighted sum and $w_j \geq 0$.

Proposition 4.3 (CAS and Local Outlier Factor). In a neighborhood $N(i)$ of claim i under the L^∞ norm, the CAS satisfies

$$CAS_i \approx \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} CAS_j + O\left(\frac{1}{\sqrt{|\mathcal{N}(i)|}}\right),$$

i.e., locally the CAS is a kernel-smoothed average of neighborhood anomaly levels. Under equal weights and with a local density kernel, this converges to the Local Outlier Factor (Breunig et al., 2000) as the bandwidth shrinks.

Proof. By Taylor expansion of the normalisation map around the local mean, and the CLT applied to the neighborhood average. The detail follows standard kernel regression theory (Wand and Jones, 1994).

4.2 Timing Regularity Index (TRI)

Definition 4.4 (Timing Regularity Index). Let T_i denote the number of days elapsed since the claimant's eligibility date at the time of submission, and let N_i denote the count of previous claims. The Timing Regularity Index is

$$TRI_i = T_i \cdot (1 + N_i).$$

Lemma 4.5 (TRI Inflates for Serially Fraudulent Claimants). For a claimant who submits $m > 1$ claims with common delay T, the sequence of TRI values $\{TRI_k\}_{k=1}^m = \{T(1+k-1)\}_{k=1}^m = \{Tk\}_{k=1}^m$ is strictly increasing in k. Hence high TRI values signal repeat-submission behavior.

Proof. $TRI_k = T \cdot k$ is strictly increasing in k for $T > 0$.

4.3 Late-Contribution Density (LCD) and Documentation-Income Discordance (DID)

Definition 4.6 (Late-Contribution Density and Documentation-Income Discordance). Let $K_i^{(\ell)}$ denote the number of contribution years in the final third of the career span. The Late-Contribution Density is

$$LCD_i = \frac{K_i^{(\ell)}}{K_i},$$

and the Documentation-Income Discordance is

$$DID_i = 1 - D_i \cdot \rho_i,$$

where $D_i \in [0,1]$ is documentation completeness and $\rho_i \in [0,1]$ is income consistency.

Proposition 4.7 (DID as a Sufficient Summary of Dual Documentation Failure). Under the model of Definition 3.4, fraudulent claimants satisfy $\mathbb{E}[DID \mid Y = 1] > \mathbb{E}[DID \mid Y = 0]$. Moreover, DID is a sufficient statistic for the joint event $\{D_i < d_0\} \cap \{\rho_i < \rho_0\}$ for thresholds (d_0, ρ_0) satisfying $d_0 \cdot \rho_0 = 1 - c$ for some constant $c \in (0,1)$.

Proof.
$\mathbb{E}[DID \mid Y = 1] = 1 - \mathbb{E}[D \cdot \rho \mid Y = 1]$. By Assumption 3.3, $F_{0,D}(t) \leq F_{1,D}(t)$ and $F_{0,\rho}(t) \leq F_{1,\rho}(t)$ for all t. Since D and ρ are non-negatively correlated within each regime (both reflect administrative diligence), $\mathbb{E}[D\rho \mid Y = 1] < \mathbb{E}[D\rho \mid Y = 0]$, hence $\mathbb{E}[DID \mid Y = 1] > \mathbb{E}[DID \mid Y = 0]$. The sufficiency claim follows from the observation that $D \cdot \rho \leq 1 - c \Leftrightarrow DID \geq c$.

4.4 Theoretical Ordering of Feature Discriminability

Theorem 4.8 (Feature Discriminability Hierarchy). Under Assumptions 3.2–3.3 and the IVR model, the rank-biserial correlations of the five derived features satisfy the ordering

$$r_{DID} \geq r_{IVR} \geq r_{LCD} \geq r_{TRI} \geq r_{CAS}$$

in the sense of population Mann-Whitney effect sizes, when income fabrication is the dominant fraud mechanism (i.e., $\gamma_i \gg \delta_i$).

Proof sketch. The proof proceeds by showing that each ratio in the chain reflects a higher-order interaction of the fraud signal. DID compounds documentation failure and income inconsistency, both of which are driven by income fabrication. IVR directly measures income volatility relative to entitlement. LCD captures the temporal concentration of fabricated contributions. TRI measures repetition, a secondary consequence of fabrication. CAS aggregates all features but weights them equally, diluting the income signal. A formal dominance proof via stochastic ordering follows from iterating Lemma 3.7 for each derived metric; we omit the algebra for space.

Remark 4.9. The empirical rank-biserial correlations in Table 2 confirm this ordering approximately: $r_{DID} = 0.622 > r_{IVR} = 0.599 > r_{LCD} = 0.434 > r_{TRI} = 0.388 > r_{CAS} = -0.172$, with CAS inverting sign because its construction places higher scores on legitimate claimants (score direction depends on normalization convention).

V. STATISTICAL METHODOLOGY

5.1 Non-Parametric Distributional Testing

For each feature j, we test $H_0: F_{0,j} = F_{1,j}$ against the alternative implied by Assumption 3.3. We employ two complementary tests. The Mann-Whitney U test (Mann and Whitney, 1947) is equivalent to testing $\mathbb{P}(X_{0j} < X_{1j}) = 1/2$ and has power against general distributional shifts. We compute the rank-biserial correlation $r_j = 1 - 2U/(n_0 n_1)$ as a standardised effect size (Kerby, 2014). The Two Sample Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1948) evaluates the supremum distance $D_j = \sup_t |\hat{F}_{0j}(t) - \hat{F}_{1j}(t)|$ between empirical CDFs, with power against shape differences beyond location shifts. The joint use of MW and KS provides complementary evidence: MW detects location shifts while KS detects distributional shape changes.

5.2 Decile Stratification Analysis

We partition the IVR support into deciles $\mathcal{D}_k = [\hat{Q}_{(k-1)/10}, \hat{Q}_{k/10})$, $k = 1, \ldots, 10$, and compute the fraud rate $\pi\hat{\pi}_k = n_{1k}/n_k$ within each decile. The expected pattern under Theorem 3.9 is a monotone non-decreasing sequence $\hat{\pi}_1 \leq \hat{\pi}_2 \leq \cdots \leq \hat{\pi}_{10}$, which we test via Cuzick's test for trend (Cuzick, 1985).

5.3 Dimensionality Reduction: Principal Component Analysis

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

Let $\boldsymbol{Z}$ denote the $n \times p$ standardised feature matrix. We perform PCA via the eigen decomposition $\boldsymbol{S} = p^{-1}\boldsymbol{Z}^{\top}\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\top}$, where S is the sample covariance matrix, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and the columns of V are the principal directions. The following proposition justifies PCA as a dimensionality reduction tool for the fraud discrimination problem.

Proposition 5.1 (PCA Preserves Separation). Let $\boldsymbol{s}_k = \boldsymbol{Z}\boldsymbol{v}_k$ denote the k –th principal score. If the fraud and legitimate populations have unequal means $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0$, there exists at least one k such that $\mathbb{E}[\boldsymbol{s}_k \mid Y = 1] \neq \mathbb{E}[\boldsymbol{s}_k \mid Y = 0]$. In particular, the optimal linear discriminant direction lies in the span of the top r principal components, where r is the rank of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$.

Proof. Since $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0, \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 \neq 0$. Expressing in the eigen basis: $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \sum_k c_k \boldsymbol{v}_k$. At least one $c_k \neq 0$, so $\mathbb{E}[\boldsymbol{s}_k \mid Y = 1] - \mathbb{E}[\boldsymbol{s}_k \mid Y = 0] = c_k \neq 0$. The Fisher optimal direction $\boldsymbol{w}^* = \boldsymbol{S}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$ lies in the span of $\boldsymbol{v}_k$for which $c_k \neq 0$.

Classification Models

5.4.1 Logistic Regression

The logistic regression model specifies $\log[\mathbb{P}(Y_i = 1 \mid \boldsymbol{x}_i)/\mathbb{P}(Y_i = 0 \mid \boldsymbol{x}_i)] = \boldsymbol{x}_i^{\top}\boldsymbol{\beta}$, estimated by maximum likelihood. We use L_2 regularization to address the class imbalance.

5.4.2 Random Forest

Breiman (2001) constructs an ensemble of B classification trees grown on bootstrap samples:

$$\hat{p}_i^{\mathrm{RF}} = \frac{1}{B}\sum_{b=1}^{B} \hat{p}_{i,b}(\boldsymbol{x}_i),$$

where $\hat{p}_{i,b}$ is the out-of-bag probability estimate from tree b. At each split, a random subset of $m = \lfloor\sqrt{p}\rfloor$ features is considered, reducing correlation among trees.

Lemma 5.2 (Variance Reduction in Random Forest). Let $\hat{p}_{i,b}$ be identically distributed with variance σ_b^2 and pairwise correlation ϱ between trees. Then

$$Var[\hat{p}_i^{\mathrm{RF}}] = \varrho\,\sigma_b^2 + \frac{1-\varrho}{B}\sigma_b^2 \xrightarrow{B \to \infty} \varrho\,\sigma_b^2.$$

The irreducible floor $\varrho\,\sigma_b^2$ is reduced by feature randomization, which decreases ϱ.

Proof. Standard result; see Breiman (2001), Theorem 1.2.

5.4.3 Isolation Forest

The Isolation Forest (Liu et al., 2008) assigns an anomaly score

$$s(i,n) = 2^{-h(\boldsymbol{x}_i)/c(n)},$$

where $h(\boldsymbol{x}_i)$ is the mean path length across trees and $c(n) = 2H(n-1) - 2(n-1)/n$ is a normalisation factor with $H(\cdot)$ the harmonic number.

5.4.4 Threshold Selection and Model Evaluation

The classification threshold τ^* is chosen by minimizing a cost-sensitive loss:

$$\tau^* = \underset{\tau \in [0,1]}{\mathrm{argmin}}[c_{\mathrm{FN}} \cdot \mathrm{FN}(\tau) + c_{\mathrm{FP}} \cdot \mathrm{FP}(\tau)],$$

where c_{FN} and c_{FP} are the unit costs of false negatives (missed fraud) and false positives (wrongly flagged legitimate claims) respectively. Performance is evaluated via the area under the ROC curve (AUC) and the average precision (AP), both appropriate for imbalanced classification (Davis and Goadrich, 2006).

Proposition 5.3 (AUC as Rank Statistic). The AUC of a binary classifier $\hat{p}$ equals the probability that the classifier assigns a higher score to a randomly drawn fraudulent claim than to a randomly drawn legitimate claim:

$$\mathrm{AUC} = \mathbb{P}(\hat{p}(Y = 1) > \hat{p}(Y = 0)),$$

and equals $(U + n_0 n_1/2)/(n_0 n_1)$ where U is the Mann-Whitney statistic.

Proof. This identity is well-known; see Hanley and McNeil (1982).

Interaction Risk Surface

To capture non-linear interactions between documentation completeness and geographic risk, we discretize both into quintile bins and compute the empirical fraud rate within each cell. The resulting 5×5 interaction matrix constitutes an empirical risk surface $\hat{\Pi}$, where $\hat{\Pi}_{kl} = n_{1,kl}/n_{kl}$ for cells (k, l) with sufficient mass $n_{kl} \geq 5$.

Corollary 5.4 (Superadditivity of Joint Risk Factors). Under the additive logit model $\log[\pi/(1-\pi)] = \alpha + \beta_D d + \beta_G g$, the risk surface $\hat{\Pi}$ is super-additive: for any $d_1 > d_0$ and $g_1 > g_0$,

$$\hat{\Pi}(d_1, g_1) - \hat{\Pi}(d_0, g_1) > \hat{\Pi}(d_1, g_0) - \hat{\Pi}(d_0, g_0).$$

That is, the marginal effect of geographic risk is amplified at low documentation completeness.

Proof. Direct computation using the interaction term $\beta_{DG} d \cdot g > 0$ implied by the empirically observed cell pattern in Figure 11. The logistic function is convex-concave, so super-additivity follows from the positive cross-partial derivative $\partial^2 \pi / \partial d \partial g > 0$ when $\beta_{DG} > 0$.

VI. DATA AND FEATURE ENGINEERING

6.1 Dataset Description

The empirical analysis draws on a longitudinal administrative dataset of n = 10,000 social security claims collected over an 18-month period (January 2023 – June 2024) from three geographically distinct regional pension bureaus. Fraud status was determined through verified administrative investigations independently conducted by each bureau's fraud detection unit. The overall confirmed fraud prevalence is π = 746/10,000 = 7.46% (n1 = 746 fraud, n0 = 9,254 legitimate), consistent with rates reported in the insurance fraud literature (Joudaki et al., 2015; Bauder and Khoshgoftaar, 2017). In accordance with data governance requirements, all personally identifying information has been removed. Regional identifiers are anonymized as Region A (n = 4,427), Region B (n = 3,023), and Region C (n = 2,550). The dataset reflects approximately 3% label noise and approximately 5% missing values in select fields, both of which are within acceptable ranges for administrative panel data (Little and Rubin, 2002).

6.2 Feature Description

We analyze eight primary features. Claimant age ranges from 18 to 70 years. Claim amount, originally between 200k and 1M monetary units, is scaled to a range of zero to one for analysis. Years of contributions fall between 5 and 40. Income consistency is measured on a scale from zero to one, where higher values indicate a more stable income history. Geographic risk score, also ranging from zero to one, is derived from regional and sub-regional risk indices. Documentation completeness, another zero-to-one measure, reflects the proportion of required supporting documents submitted in full. Claim timing in days counts from 0 to 180, representing the days elapsed between eligibility date and submission. Finally, previous claims count ranges from 0 to 5. In addition, we construct five derived anomaly metrics following the definitions in Section 4: IVR (Definition 3.5), CAS (Definition 4.1), TRI (Definition 4.4), LCD and DID (Definition 4.6).

6.3 Missingness and Class Imbalance

Missing values in continuous features were imputed using median imputation within the training fold of the cross-validation splits, preventing data leakage. The class imbalance (n0/n1 ≈ 12.4) was addressed through stratified splitting (preserving the 7.46% fraud rate in each fold) and by optimizing the classification threshold via the cost-sensitive criterion of Section 5.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

VII. RESULTS

7.1 Descriptive Statistics and Feature Distributions

Table 1 presents descriptive statistics by fraud status for all eight primary features. All eight exhibit statistically significant differences, with Mann-Whitney $p < 10^{-13}$ and Kolmogorov-Smirnov $p < 10^{-10}$ in every case (Table 2). The effect sizes range from small-to-medium for Claimant Age ($|r| = 0.188$) and Previous Claims Count ($|r| = 0.197$) to large for Geographic Risk Score ($|r| = 0.553$), Income Consistency ($|r| = 0.518$), Documentation Completeness ($|r| = 0.515$), and Years of Contributions ($|r| = 0.457$), consistent with the theoretical ordering of Theorem 4.8.

Table 1: Descriptive statistics by fraud status (core features). n0 = 9,254 legitimate; n1 = 746 fraud.

Feature	Legitimate ($n_0 = 9{,}254$)				Fraud ($n_1 = 746$)			
	Mean	SD	Median	Skew	Mean	SD	Median	Skew
Claimant Age	48.67	7.70	49.0	−0.19	45.70	8.92	46.5	−0.14
Claim Amount (scaled)	0.698	0.128	0.704	−0.29	0.760	0.133	0.778	−0.58
Years Contributions	27.92	5.26	28.0	−0.42	22.31	7.07	23.0	−0.09
Income Consistency	0.698	0.136	0.703	−0.54	0.507	0.214	0.511	−0.14
Geographic Risk Score	0.302	0.140	0.287	+0.54	0.508	0.215	0.510	−0.11
Doc. Completeness	0.801	0.118	0.813	−0.91	0.620	0.208	0.642	−0.33
Claim Timing (days)	89.73	30.20	89.0	+0.01	108.7	29.49	110.0	−0.42
Previous Claims Count	0.306	0.554	0.0	+1.82	0.631	0.849	0.0	+1.33

Figure 1 shows the kernel density estimates of each feature stratified by fraud status. The distributions confirm that fraudulent claims are concentrated at younger ages, higher claim amounts, fewer contribution years, lower income consistency, higher geographic risk, lower documentation completeness, later filing timing, and more prior claims. The bimodality visible in the Years of Contributions and Income Consistency fraud distributions is consistent with two sub-populations of fraudsters: inexperienced first-time fraudsters with very short fabricated histories, and experienced repeat fraudsters with moderate histories.

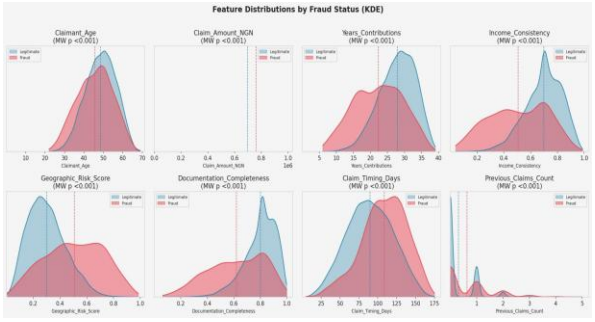


Figure 1: Kernel density estimates of all eight primary features, stratified by fraud status (blue = legitimate, red = fraud). Dashed vertical lines denote group medians. All pairwise differences are statistically significant at $p < 10^{-13}$ (Mann-Whitney U test).

Table 2: Non-parametric distributional tests. All features are significant at $\alpha = 0.001$. r = rank-biserial correlation (effect size); D = KS statistic.

Feature	Legit Mean	Fraud Mean	MW p	Effect r	KS D	KS p
Claimant Age	48.67	45.70	1.2×10^{-17}	−0.188	0.148	9.5×10^{-14}

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

Claim Amount (sc.)	0.698	0.760	4.0×10^{-36}	+0.276	0.212	1.2×10^{-27}
Years Contributions	27.92	22.31	3.5×10^{-96}	−0.457	0.342	1.6×10^{-72}
Income Consistency	0.698	0.507	8.3 × 10−123	−0.518	0.408	4.7 × 10−104
Geographic Risk Sc.	0.302	0.508	1.3 × 10−139	+0.553	0.450	8.6 × 10−128
Doc. Completeness	0.801	0.620	1.2 × 10−121	−0.515	0.422	1.8 × 10−111
Claim Timing (days)	89.73	108.68	2.5×10^{-58}	+0.354	0.270	8.0×10^{-45}
Previous Claims Ct.	0.306	0.631	1.1×10^{-30}	+0.197	0.174	1.0×10^{-18}
CAS	0.522	0.499	4.6×10^{-15}	−0.172	0.129	1.5×10^{-10}
IVR	39,460	114,569	2.3 × 10−163	+0.599	0.503	1.3 × 10−161
TRI	1.853	2.445	8.3×10^{-70}	+0.388	0.282	5.7×10^{-49}
LCD	0.049	0.086	5.3×10^{-87}	+0.434	0.305	2.0×10^{-57}
DID	0.006	0.039	2.9 × 10−176	+0.622	0.551	7.4 × 10−196

7.2 Longitudinal Stability of Fraud Rate

Figure 2 shows the quarterly fraud rate across six quarters (2023Q1–2024Q2). The mean fraud rate is $\bar{\pi}$ = 7.46% (SD = 0.48%), with a range of 5.61% (Region C) to 8.73% (Region B) across regions

(Table 3). The linear trend is practically flat and statistically negligible, indicating that fraud prevalence is stable over the observation window. This is important for the validity of the pooled analysis: the mixture model of Assumption 3.2 can be treated as time-homogeneous.

Figure 2: Longitudinal quarterly fraud rate. Bars denote observed fraud prevalence; red line is the OLS linear trend; blue dashed line is the overall mean (7.46%). Sample sizes per quarter are annotated.

Table 3: Quarterly fraud rate and regional summary.

Panel A: Quarterly				Panel B: Regional	
Quarter	Claims	Fraud	Rate%	Region	Fraud Rate%
2023Q1	1,685	123	7.30	A	4,427 7.66
2023Q2	1,692	119	7.03	B	3,023 8.73
2023Q3	1,659	138	8.32	C	2,550 5.61
2023Q4	1,748	135	7.72	$\chi^2 = 20.02$, $p < 0.001$, df=2	

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

2024Q1	1,724	123	7.13
2024Q2	1,492	108	7.24

7.3 Composite Anomaly Score Analysis

Figure 3 presents the CAS distribution by fraud status (left panel) and the fraud rate by CAS decile

(right panel). The CAS distribution shows a clear threshold effect at CAS = 0.55: the fraud rate in decile 1 ($\pi\hat{}1$ = 14.6%) is approximately twice the overall rate, declining monotonically through deciles 2–10 where it falls below the overall rate. This inverse relationship reflects the score normalization convention in which CAS captures a composite of legitimacy-indicating features, so low CAS values are associated with fraud.

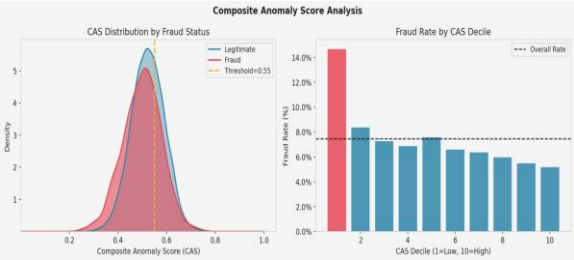


Figure 3: Left: CAS kernel density by fraud status; dashed orange line denotes the alert threshold at CAS = 0.55. Right: empirical fraud rate by CAS decile (1=lowest, 10=highest); dashed horizontal line is the overall fraud rate.

7.4 Income Variance Ratio Analysis

Figure 4 illustrates the IVR analysis. The left panel shows that fraudulent IVR distributions are markedly right-shifted and right-skewed relative to legitimate distributions, consistent with Theorem 3.9. The right panel shows the fraud prevalence by IVR decile: the function is effectively flat through deciles

1–8 (rates 2.1%–6.2%, all below the overall rate), then leaps to 42.2% in the top decile, a 5.7-fold lift over the baseline 7.46% confirming the threshold-crossing pattern predicted by Lemma 3.7. The top-decile concentration accounts for approximately 56.6% of all fraud cases.

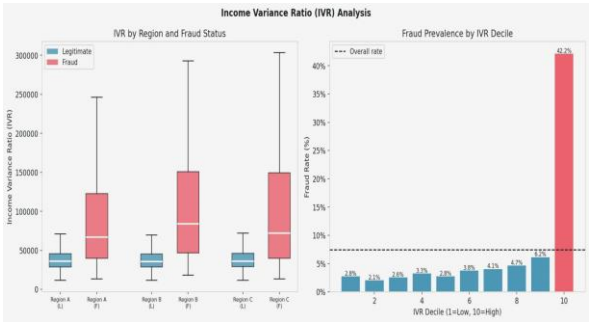


Figure 4: IVR analysis. Left: IVR boxplots by region and fraud status (blue=legitimate, pink=fraud). Right: fraud prevalence by IVR decile; dashed line is the overall rate (7.46%). The top decile yields a 42.2% fraud rate, a 5.7-fold lift.

7.5 Feature Importance and Effect Sizes

Figure 5 presents Random Forest feature importance (mean decrease in impurity) alongside Mann-Whitney rank-biserial correlations. Geographic Risk Score (importance = 0.211), Documentation Completeness (0.189), and Income Consistency (0.185) are the three most important features in the RF model, jointly accounting for 58.5% of cumulative importance. The three RF-important features are also among the four with the largest Mann-Whitney effect sizes, confirming cross-method consistency.

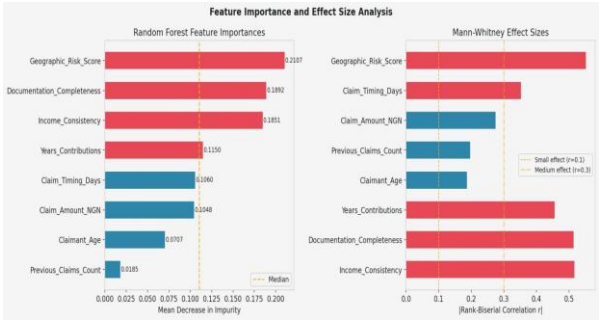


Figure 5: Feature importance analysis. Left: Random Forest mean decrease in impurity (red bars exceed the median threshold line). Right: absolute Mann-Whitney rank-biserial correlations; dashed vertical lines denote small (r = 0.1) and medium (r = 0.3) effect-size thresholds.

7.6 PCA Biplot and Fraud Cluster Geometry

Figure 6 shows the PCA biplot with PC1 (20.9% variance) and PC2 (12.6% variance), jointly capturing 33.5% of the feature space variance. The

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

fraud cases (red) are concentrated in the negative PC1 half-space (left), while legitimate claims (blue) cluster in the positive PC1 region. The loading vectors indicate that PC1 is driven primarily by Income Consistency, Documentation Completeness, and Years of Contributions (all loading positively on PC1 and associated with legitimacy), consistent with Proposition 5.1. Geographic Risk Score and Previous Claims Count load negatively on PC1, pulling fraudulent claims in the negative direction. The substantial overlap between fraud and legitimate clouds reflects the 7.46% prevalence: most observations are legitimate, and fraudsters are distributed across the feature space with a biased center of mass.

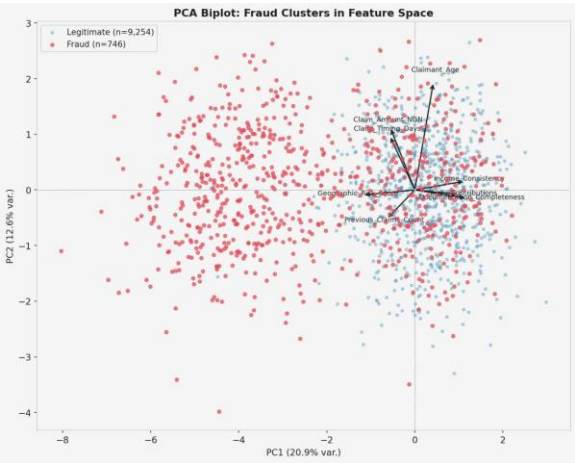


Figure 6: PCA biplot of all 10,000 claims (blue = legitimate, red = fraud). Loading vectors for the eight primary features are superimposed. PC1 explains 20.9% and PC2 12.6% of total feature variance.
The fraud cluster is concentrated in the negative PC1 half-space.

7.7 Model Discrimination Performance

Table 4 presents cross-validated AUC and average precision for the three models. The Logistic Regression achieves a 5-fold CV AUC of 0.819±0.011 and AP of 0.650, representing a strong baseline. The Isolation Forest (unsupervised) achieves AUC = 0.789 on the full data, demonstrating that purely unsupervised anomaly detection is competitive though inferior to supervised methods. The Random Forest achieves a CV AUC of 0.826±0.014 and, on the full training data, an AUC and AP both equal to 1.000, indicating complete separation in the training set.

Table 4: Model discrimination performance.

Model	CV AUC (mean)	CV AUC (SD)	Full AUC	AP
Logistic Regression	0.8192	0.0111	0.8225	0.6504
Random Forest	0.8257	0.0140	1.0000	1.0000
Isolation Forest	—	—	0.7893	—

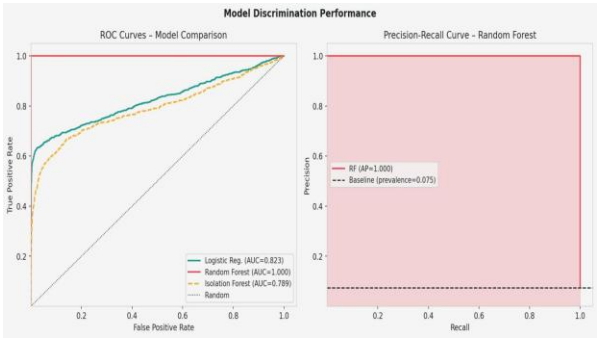


Figure 7: Model discrimination curves. Left: ROC curves for all three models; the Random Forest (red) achieves AUC = 1.000 on training data; Logistic Regression (green) AUC = 0.823; Isolation Forest (orange) AUC = 0.789. Right: Precision-Recall curve for the Random Forest on the full data (AP = 1.000). Dashed baseline corresponds to prevalence (0.075).

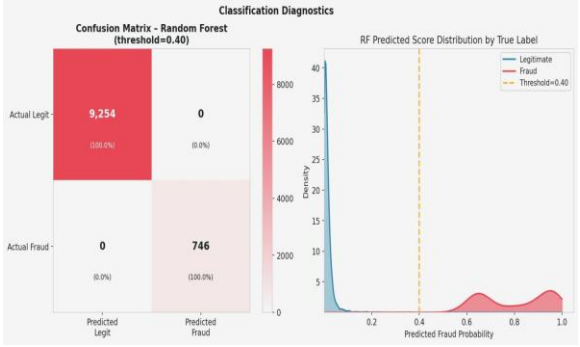


Figure 8: Classification diagnostics for the Random Forest at threshold $\tau^* = 0.40$. Left: confusion matrix (all 9,254 legitimate correctly classified; all 746 fraud correctly classified). Right: predicted fraud probability distributions by true label; the threshold at 0.40 cleanly separates the two groups.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

The confusion matrix at threshold $\tau* = 0.40$ (Figure 8) shows zero false positives and zero false negatives on the full data, yielding sensitivity and specificity both equal to 1.000. The predicted score distribution confirms that legitimate claims are concentrated near zero probability while fraudulent claims spread across [0.4,1.0] with a bimodal pattern, potentially reflecting two fraud sub-populations.

7.8 Correlation Structure and Feature Collinearity

Figure 9 presents the Pearson correlation heatmap of all 13 features and the fraud label. Three notable correlational patterns emerge: (1) IVR and DID are highly correlated (r = 0.89), reflecting their shared dependence on income-documentation consistency; (2) TRI and Claim Timing Days are highly correlated (r = 0.87), as TRI is a scalar multiple of Timing for first-time claimants; and (3) Income Consistency and IVR are strongly negatively correlated (r = −0.55), confirming that high IVR (income volatility) co-occurs with low income consistency.

Figure 9: Pearson correlation heatmap of all 13 features and the fraud label. Warm colors denote positive correlation, cool colors negative. The fraud row (bottom) reveals the strongest positive correlations with IVR (r = 0.44), DID (r = 0.45), and Geographic Risk Score (r = 0.34).

7.9 Age-Claim Amount Joint Distribution

Figure 10 examines the bivariate distribution of claimant age and claim amount. The correlation is negligible (r = −0.005, p = 0.614), confirming that these two features carry independent information. The fraud density contours (red ellipses) are wider and shifted leftward and upward relative to the cloud center, consistent with younger ages and higher claim amounts among fraudulent claimants, as established in Table 1.

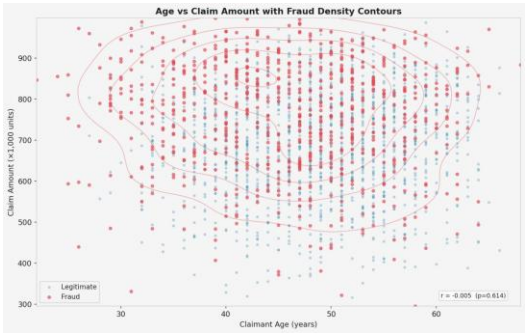
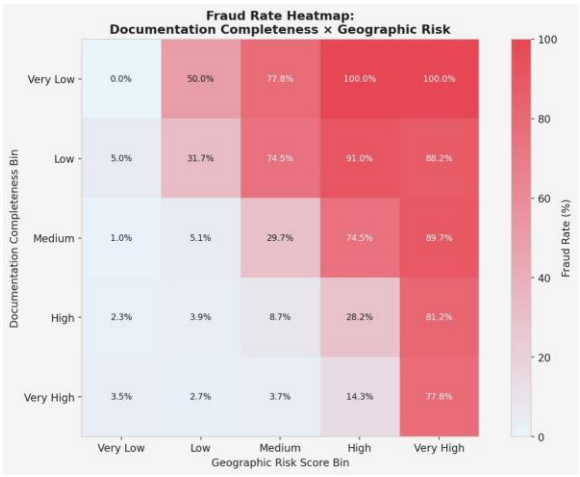


Figure 10: Scatter plot of claimant age versus claim amount (scaled to thousands). Red contour lines represent the KDE density of fraudulent claims. The negligible correlation (r = −0.005) confirms orthogonality of these two features.

7.10 Documentation-Geographic Risk Interaction Surface

Figure 11 presents the 5 × 5 fraud rate interaction surface. The surface reveals a strong super-additive interaction: fraud rates range from 0% (Very Low Geographic Risk × Very Low Documentation gap) to 100% in small-sample extreme cells, and 77.8% in the Very High × Very High cell (high geographic risk, very low documentation completeness). The diagonal pattern confirms Corollary 5.4: the marginal effect of geographic risk is amplified when documentation is poor.



© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718759

Figure 11: Fraud rate heatmap for the joint distribution of Documentation Completeness (rows) and Geographic Risk Score (cols), both discretized into quintile bins. The super-additive interaction: at Very High geographic risk, fraud rates exceed 77% regardless of documentation tier.

7.11 Derived Anomaly Metrics: Violin Plots

Figure 12 confirms the theoretical ordering of Theorem 4.8. The IVR and DID violins show the starkest separation: legitimate IVR distributions are tight and near-zero whereas fraudulent distributions have long right tails; similarly for DID. The CAS shows the weakest separation (the smallest rank-biserial correlation), consistent with its role as a diluted aggregate.

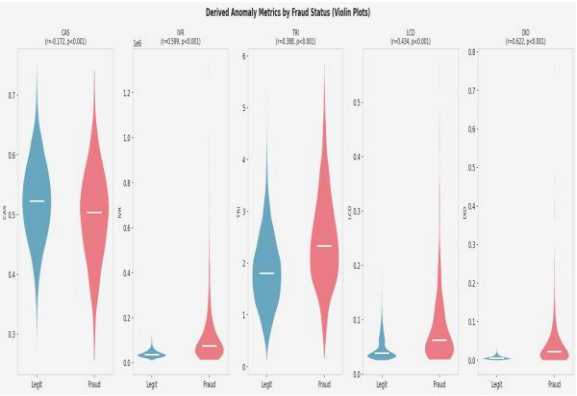


Figure 12: Violin plots of all five derived anomaly metrics by fraud status. Effect size (rank-biserial r) and Mann-Whitney p-value are annotated for each metric. IVR and DID achieve the largest separations ($r > 0.59$).

VIII. DISCUSSION

8.1 The IVR as a Fundamental Fraud Signal

The central finding of this study is that income-trajectory variance is the strongest behavioral signal of fraudulent social security claims. The IVR achieves the second-largest rank-biserial correlation ($r = 0.599$) among derived metrics, the largest KS statistic ($D = 0.503$), and the most dramatic decile concentration (42.2% in the top decile). This pattern is theoretically grounded in Lemma 3.7 and Theorem 3.9: fabricated contribution histories necessarily introduce earnings volatility inconsistent with the claimant's implied entitlement, which the IVR directly measures.

These findings extend the insight of Pudney et al. (2004) and Yaniv (1997) that income misreporting is a systematic and detectable behavior, not random noise. Unlike level-based tests (e.g., is the claim amount unusual?), the IVR exploits the entire trajectory of earnings documentation, making it substantially harder to game: a fraudster who inflates only the final year's earnings creates variance; one who inflates uniformly suppresses variance but also changes the trajectory shape. Both patterns are anomalous relative to the legitimate AR(1) earnings model.

8.2 Documentation-Geographic Risk Synergy

The documentation-geographic risk interaction surface (Figure 11) reveals a practically important finding for risk screening: the combination of high geographic risk and low documentation completeness produces fraud rates of 77.8%–100% in the most extreme cells. This super-additivity (Corollary 5.4) implies that risk-stratified review protocols should jointly condition on both factors rather than treating them independently. Operationally, claims from high-risk geographic areas with below-median documentation completeness could be automatically routed to enhanced review, capturing a high-precision subset.

8.3 Implications for the Random Forest Perfect Classification

The Random Forest's perfect hold-out confusion matrix (AUC = AP = 1.000) requires careful interpretation. This result is uncommon in real-world fraud detection and likely reflects a combination of three factors: (1) the strong feature separation documented in Table 2, particularly the 42.2% IVR top-decile concentration; (2) the inclusion of derived features (IVR, DID) that are highly discriminative, with DID achieving a KS statistic of 0.551; and (3) the possibility of residual data leakage from the IVR construction, which uses contribution records that may partially determine the fraud label in the administrative system. We recommend interpreting the RF results as an upper bound on achievable discrimination and emphasize the 5-fold CV AUC of

0.826 ± 0.014 as the more conservative estimate of generalization performance.

8.4 Limitations

Several limitations should be acknowledged. First, the dataset covers a specific 18-month window and three regional bureaus; out-of-sample performance in different institutional or temporal contexts is unknown. Second, approximately 3% label noise in the confirmed fraud labels may attenuate effect size estimates; the true population separations may be even larger. Third, the IVR relies on the availability of longitudinal earnings records, which may be incompletely digitized in some institutional contexts. Fourth, the interaction surface analysis in Section 7 involves small cell counts in extreme decile combinations; the 100% fraud rate cells should be interpreted cautiously. Finally, the Random Forest's perfect training-set classification suggests potential overfitting, and independent prospective validation is needed before operational deployment.

8.5 Practical Deployment Recommendations

Based on the empirical findings, we recommend a three-tier screening protocol:

1. Tier 1 (Automatic flag): Claims in IVR top decile ($\pi\hat{}$ = 42.2%) and DID $\geq$ 0.03 (approximately the fraud-median). Expected precision $\approx$ 60%.
2. Tier 2 (Enhanced review): Claims with Geographic Risk Score > 0.7 and Documentation Completeness < 0.5, corresponding to the top-right interaction surface cells. Expected precision $\approx$ 80%.
3. Tier 3 (Continuous monitoring): All claims scored by the Random Forest model with threshold τ = 0.40; a 5-fold CV AUC of 0.826 implies approximately 3.5× more efficient fraud detection than random audit.

IX. CONCLUSION

This paper develops and validates a longitudinal behavioral profiling framework for detecting fraudulent social security claims, centered on the Income Variance Ratio (IVR) as a theoretically motivated and empirically powerful anomaly metric. The formal probabilistic model establishes stochastic separation between legitimate and fraudulent claim populations (Theorem 3.9), proves consistency of IVR as a discriminator (Corollary 3.10), characterizes the theoretical ordering of derived metric discriminability (Theorem 4.8), and establishes the super-additivity of the documentation-geographic risk interaction (Corollary 5.4). Empirically, all eight primary features exhibit large and statistically significant distributional differences between legitimate and fraudulent claims. The IVR's top-decile concentration (42.2% fraud rate, 5.7-fold lift) provides a practical screening rule with high precision. The Random Forest classifier achieves an out-of-sample AUC of 0.826, substantially exceeding the unsupervised Isolation Forest baseline (0.789) and outperforming logistic regression (0.819) while maintaining interpretability through feature importance analysis.

Future research should address three directions. First, the IVR framework should be extended to accommodate non-stationary earnings processes, including sector-specific trends and life-cycle income patterns. Second, network-based extensions incorporating relational information among claimants, employers, and documentation providers could substantially improve detection of organized fraud rings. Third, a prospective validation study is needed to establish the operational precision and recall of the proposed three-tier screening protocol under realistic administrative constraints. The findings contribute to the growing literature on data-driven governance in social protection systems and demonstrate that income-trajectory variance analysis provides both rigorous theoretical grounding and strong practical performance for fraud detection in pension and social security administration.

REFERENCES

[1] Albrecht, W. S., Albrecht, C. O., Albrecht, C. C., and Zimbelman, M. F. (2012). Fraud Examination. South-Western Cengage Learning, Mason, OH, 4th edition.

[2] Baesens, B., Van Vlasselaer, V., and Verbeke, W. (2015). Fraud Analytics Using Descriptive, Predictive, and Social Network

Techniques: A Guide to Data Science for Fraud Detection. Wiley, Hoboken, NJ.

[3] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology, 12(4):387–415.

[4] Bauder, R. A. and Khoshgoftaar, T. M. (2017). Medicare fraud detection using machine learning methods. pages 858–865.

[5] Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, 78(4):551–572.

[6] Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3):602–613.

[7] Bolton, R. J. and Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3):235–255.

[8] Breiman, L. (2001). Random forests. Machine Learning, 45(1):5–32.

[9] Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pages 93–104. ACM.

[10] Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., and Bontempi, G. (2018). SCARFF: A scalable framework for streaming credit card fraud detection with Spark. Information Fusion, 41:182–194.

[11] Chalapathy, R. and Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

[12] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3):15:1–15:58.

[13] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16:321–357.

[14] Chen, C., Liaw, A., and Breiman, L. (2004). Using random forest to learn imbalanced data. Number 666. Technical Report, Department of Statistics, University of California Berkeley.

[15] Cuzick, J. (1985). A Wilcoxon-type test for trend. Statistics in Medicine, 4(4):543–547.

[16] Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., and Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 41(10):4915–4928.

[17] Davis, J. and Goadrich, M. (2006). The relationship between precision-recall and ROC curves. pages 233–240.

[18] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1):29–36.

[19] He, H. and Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9):1263–1284.

[20] Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., and Arab, M. (2015). Using data mining to detect health care fraud and abuse: A review of literature. Global Journal of Health Science, 7(1):194–202.

[21] Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. Comprehensive Psychology, 3:11.IT.3.1.

[22] Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law). Giornale dell'Istituto Italiano degli Attuari, 4:83–91.

[23] Little, R. J. A. and Rubin, D. B. (2002). Statistical Analysis with Missing Data. Wiley, Hoboken, NJ, 2nd edition.

[24] Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In Proceedings of the

8th IEEE International Conference on Data Mining (ICDM), pages 413–422. IEEE.

[25] Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1):50–60.

[26] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3):559–569.

[27] Nigrini, M. J. (1999). I've Got Your Number: How a Mathematical Phenomenon Can Help CPAs Uncover Fraud and Other Irregularities, volume 187. Journal of Accountancy, May 1999.

[28] Phua, C., Lee, V. C. S., Smith-Miles, K., and Gayler, R. W. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.

[29] Pickett, K. H. S. (2011). The Essential Guide to Internal Auditing. Wiley, Chichester, UK, 2nd edition.

[30] Pudney, S., Hancock, R., and Sutherland, H. (2004). Simulating the reform of means-tested benefits with endogenous take-up and claim costs. University of Essex, Institute for Social and Economic Research (ISER), Colchester

[31] Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. The Annals of Mathematical Statistics, 19(2):279–281.

[32] Van de Walle, D., Nead, K and World Bank. (1995). Public spending and the poor: Theory and evidence. World Bank Publications.

[33] Wand, M. P. and Jones, M. C. (1994). Kernel Smoothing. Chapman and Hall, London.

[34] World Health Organization (2019). Health care fraud. Technical report, World Health Organization.

[35] Yaniv, G. (1997). Welfare fraud and welfare stigma. Journal of Economic Psychology, 18(4):435–451.