Quality Certification under Uncertainty: An Analysis of Wine Competition Ratings

Gianni De Nicolò
Johns Hopkins University
Magalie Dubois
Burgundy School of Business

ASSA/AAEA January 2026

Motivation

- Wine competitions certify wine quality via medals.
- Ratings influence consumer choice and producer strategies
- Current rating protocols generate two sources of uncertainty:
  - **rating risk**: Variation in jury assignments and score aggregation within competitions force producers to submit samples to multiple wine competitions, increasing their marketing costs.
  - **competition risk**: Inconsistent evaluation standards across competitions increases consumers' search costs and distorts producers' decisions
- **Reducing rating and competition risks may be welfare improving by increasing the quality of certification**

This paper

- We design a rating system aimed at reducing rating risk based on two features:
  - (a) Standardization of judges' scores, and
  - (b) Partitioning of scores into ranked, disjoint, quality-equivalent rating classes.

- We use a model of multiple certifiers to show that standardization of rating protocols reduces competition risk and improves welfare.

Contributions to the literature

- *Expert ratings and ratings aggregation in wine markets.*
  - We extend De Nicolò (2024) rating system to multiple juries, introducing statistical significance of score differences to generate rating classes—a feature not addressed in prior literature.
- *Theory of certification and quality disclosure.*
  - We extend the model by Hopenaym and Saedi (2025) to show the welfare properties of the standardization of rating protocols.

Global Overview

- 600+ competitions annually.
- Europe and North America dominate.

Figure: **Wine Competitions Worldwide (GWRM 2020-2021)**

Inconsistent Rating Protocols

Table: **Top five International Competitions: Medals and Score Ranges**

Competition	**Avg. Entries**	**Medals % of entries**	**Medal Categories and Score Ranges**
DWWA	18,500	82%	Bronze: 86–89, Silver: 90–94, Gold: 95–96, Platinum: 97–100
MV	12,000	30%	Silver: 85-89, Gold: 90-94, Grand Gold: 95-100
IWC	7,000	71%	Bronze: 85–89, Silver: 90–94, Gold: 95–100
IWSC	12,000	83%	Bronze: 75–79, Silver: 80–89, Gold: 90–100
CMB	10,000	30%	Silver: 85-87.9, Gold: 88–91.9, Gran Gold: 92-100

Table: **Top five International Competitions: Scoring Methodology**

Competition	Scoring Methodology
DWWA	Wines are blind tasted in regional flights with contextual information (grape, vintage, price band). Scores are finalized through panel discussion. Silver and Gold medals are re-tasted by senior judges and Co-Chairs.
MV	Wines are blind tasted. The rating protocol follows OIV (2025).
IWC	Wines are blind tasted in flights grouped by style and region. Co-Chairs re-taste all medal candidates.
IWSC	Wines are blind tasted. Scores are discussed in panels.
CMB	Wines are blind tasted and Wines are grouped by type. The rating protocol follows OIV (2025).

Wine Competition Scoring

- There are P juries ($p \in \{1, .., P\}$), each composed by N^p judges ($i \in \{1, .., N^p\}$), who evaluate M^p wines ($j \in \{1, .., M^p\}$)

- Judges evaluate wines according to a set of quality "factors" indexed by $q \in \{1, .., Q\}$ on a numerical rating scale.

- The score of a wine of judge i in jury p is $X_{ij}(p) = \sum_{q=1}^{Q} w_q X_{ij}^q(p)$,

- The aggregate score of wine j of jury p is $X_j(p) = \sum_{i=1}^{N^p} \sum_{q=1}^{Q} w_q X_{ij}^q(p)$

Two-step Score Standardization

- (Step 1) The *Z-score* of judge i of wine j is

$$Z_{ij}(p) = (X_{ij}(p) - \mu_i(p))\sigma_i^{-1}(p)$$

- (Step 2) The (double standardized) score of judge i of wine j in jury p is:

$$\tilde{Z}_{ij}(p) = \mu_R + \sigma_R Z_{ij}(p)$$

- μ_R and σ_R are the mean and standard deviation of the raw aggregate scores, called the *R distribution*.

The Standardized Score Distribution (S distribution)

- Mean and standard deviation of the S distribution are μ_S and σ_S respectively
- By construction,

$$\mu_S = \mu_R$$

- For a sufficiently large N

$$\sigma_S < \sigma_R$$

- Rating risk is reduced by standardization

Ranked Disjoint Rating Classes (1)

- For each jury, we use ANOVA to compute the *Fisher Least Significant Difference* (FLSD(p)) at a 5% significance level.
- The estimated FLSD(p) is used as an *indivisible unit of account*, or "currency".
- The **Quality Value (QV)** of a wine j in jury p's *currency* is:

$$QV(Z(p)_j) = int(\frac{Z(p)_j}{FLSD(p)}) \tag{1}$$

where the *int* operator transforms QV in an integer number.

- **The QV automatically delivers ranked quality equivalent classes of the wine evaluated by jury p**

Ranked Disjoint Rating Classes (2)

- QVs differ across juries, as reflected in different values of $FLSD(p)$.
- We need to convert juries' QVs in just one currency.
- Using different $FLSD$ changes is the granularity of the rating classes: the larger the $FLSD$, the less granular is the distribution of rating classes.
- *How to choose the FLSD?*

QV targeting

- The FLSD is chosen by *QV-targeting* .
- First, we use the mean FLSD across juries
- Medal categories determined subject to the constraint that the number of winning wines does not exceed the maximum percent τ of submissions.
- If such constraint is violated using the FLSD mean, then the FLSD is recomputed to derive the desired number of rating classes.

Application: Citadelles du Vin (CdV) data

- Citadelles du Vin is an International Vine and Wine Organization (OIV) sponsored wine competition
- The OIV rules allow to award only 30% of the samples and state that no jury should taste more than 45 wines per day of competition.
- The current study uses a dataset provided by Citadelles du Vin, including 123 white wines and 266 red wines tasted during the 2022 competition.

A Model of Certification under Uncertainty

- *Wine producers*. A unit mass of wine producers indexed by wine quality z distributed according to unobserved cdf $F(z)$.
- *Wine consumers*. A mass M of consumers who decide whether to purchase a wine based on preferences $U = z + \theta - p$, where $\theta \geq 0$ is a taste parameter.
- *Wine competitions*. n wine competitions deliver a score distribution $G_i(z), i = 1, ., n$. The information is represented by the **mixture** of distributions $G(z) = n^{-1} \sum_{i=1}^{n} G_i(z)$.
- *Equilibrium and welfare*. The competitive equilibrium delivers total (consumer + producer) surplus as a value function $TS(G(z))$

Welfare Implications

- *If rating quality is not distinguishable by consumers and producers, then total surplus is reduced relative to the most informative rating system.*

$$TS(G(z)) \leq \max\{TS(G_1(z)), TS(G_2(z)), ..., TS(G_n(z))\}$$

- **Implication:** higher search costs and lower producer expected profits.

- *If rating quality is distinguishable by consumers and producers, then total surplus is increased.*

$$TS(G(z)) \geq \max\{TS(G_1(z)), TS(G_2(z)), ..., TS(G_n(z))\} \quad (2)$$

- **Implication:** More precise information (i.e., less uncertainty) improves welfare

Policy Implications

- Perfect discrimination and quality assessment of rating protocols may be unattainable
- However, *standardization of rating protocols* can improve the quality of information
- Recommendations
  - *Disclose the percentage of wines eligible for medals.*
  - *Implement and disclose a standardized method for judges' scores.*
  - *Adopt and disclose a method determining rating classes based on statistically significant score differences.*

Conclusions

- To reduce rating risk, we propose a statistically grounded rating system that standardizes judges' scores and partitions wines into quality-equivalent classes based on statistically significant score differences.

- Using a model of intermediary certifiers, we show that standardization reduces competition risk, enhancing both consumer and producer welfare.