

# **Discrepancies in Wine Trade Data: A serious cause for concern?**

---

Franziska Thiemann,  
Stefan Guettler and Rolf A.E. Mueller

Department of Agricultural Economics  
Christian-Albrechts-University at Kiel, Germany

# Errors and discrepancies in trade statistics

---

- Statistics are man made artifacts
- Errors are common
  - random errors
  - systematic errors (biases)
- Quality programs of statistical offices
- Discrepancies in trade statistics
  - Morgenstern (1950)
  - Yeats (1990)
  - Rozanski and Yeats (1994)

# Causes for data errors and discrepancies

---

- Errors in statistical procedures
  - Misclassification
  - Data collection
  - Data handling and processing
- Different measurement points
  - Exchange rates
  - Transit issues
- Falsification
  - Criminal activities
  - Intentional data falsification by government and companies

# Outline

---

- Country and data selection
- Test I – Inspection of single series – Benford's Law
- Test II – Comparison of different data sources
- Test III – Comparison of pairwise trade data
- Conclusion

# Selected countries

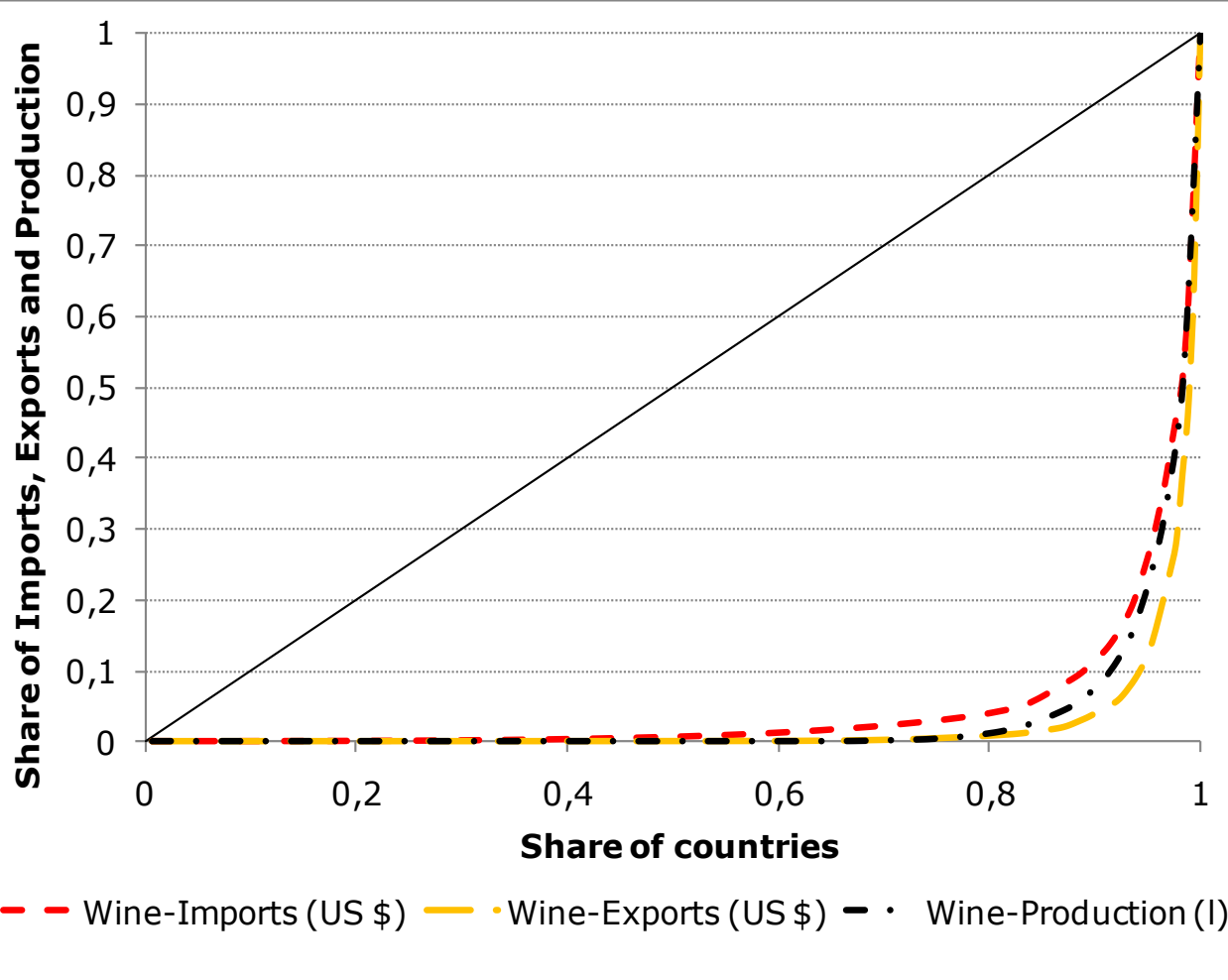
## Lorenz curves of wine production, exports and imports

**n=177**  
**yearly average 2004-2006**

Selected countries:

AUS, BEL, CAN, CHL, FRA,  
DEU, ITA, JPN, NLD, ESP,  
CHE, GBR, USA

Share of worlds exports: 85,6%  
Share of worlds imports: 77,6%



Data: OIV (n.d.), UN Comtrade (2010)

# Data sources

---

Selected Commodity Codes (SITC Rev. 3)

SITC-code 1121: Wine of fresh grapes (including fortified wine); grape must in fermentation or with fermentation arrested

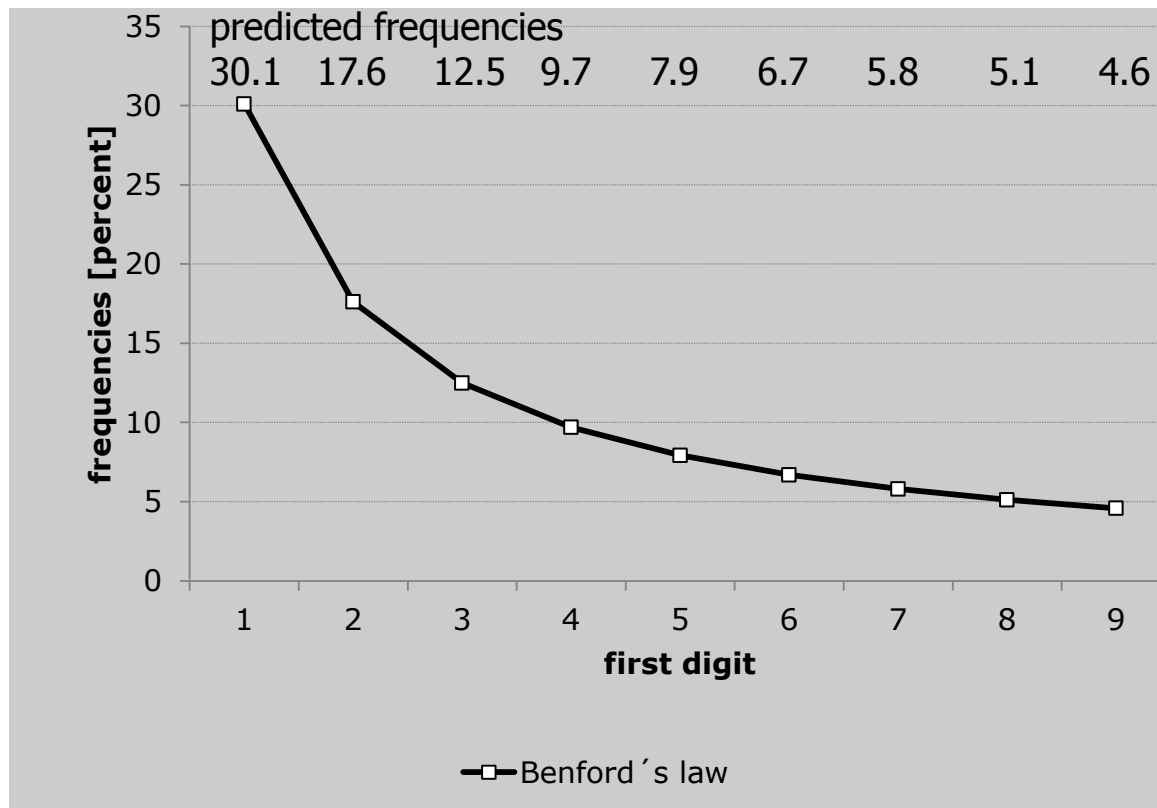
SITC-code 11217: Wine of fresh grapes (other than sparkling wine); grape must with fermentation arrested by the addition of alcohol

- UN Comtrade
- OECD
- OIV
  
- national statistics:
  - **ABS** (Australian Bureau of Statistics)
  - **USDA-FAS** (United States Department of Agriculture; Foreign Agricultural Service)
  - Eurostat

# Test I - Analysis of first digit

Benford's law:  $\text{Prob}(\text{first significant digit}=d) = \log_{10}(1+1/d)$

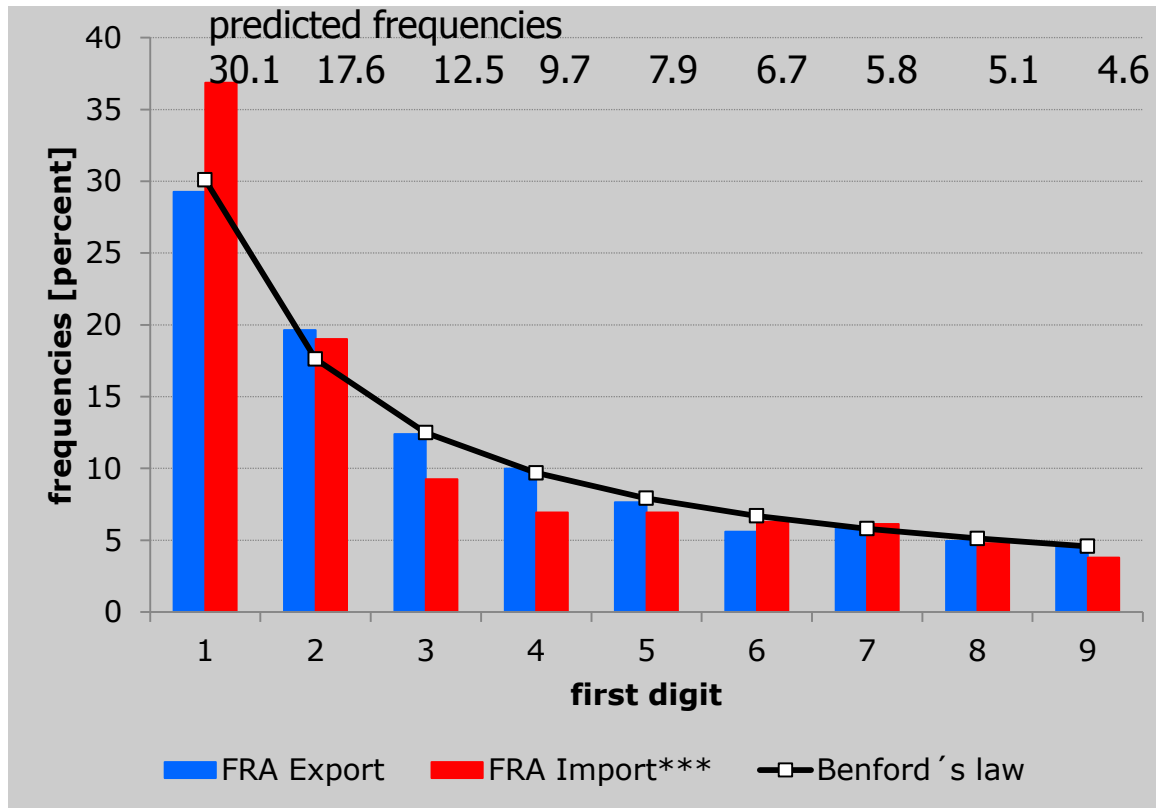
## Benford's law



# Test I - Analysis of first digit

Benford's law:  $\text{Prob}(\text{first significant digit}=d) = \log_{10}(1+1/d)$

## Benford's law and France's wine trade (SITC 11217, 2000-2008)



Conspicuous countries

Export : CHE, CHL

Import: DEU, FRA, JPN

( SITC 11217, 2000-2008)



# Results Test II: Comparison of alternative data sources

Country	Export				Import			
	UN Comtrade [1,000 hl]	Ratio			UN Comtrade [1,000 hl]	Ratio		
		OIV / UN C*	OECD / UN C*	national Statistics / UN C*		OIV / UN C*	OECD / UN C*	national Statistics / UN C*
ITA	17,123	1.07	1.02	0.91 <sup>c</sup>	1,382	1.06	1.02	1.04 <sup>c</sup>
FRA	13,171	1.12	1.28	0.96 <sup>c</sup>	5,257	1.01	0.52	0.97 <sup>c</sup>
ESP	7,892	1.82	1.00	1.57 <sup>c</sup>	70	5.70	1.00	7.35 <sup>c</sup>
AUS	7,466	1.02	0.94	0.99 <sup>a</sup>	232	1.28	1.87	0.97 <sup>a</sup>
USA	3,699	1.02	1.00	0.98 <sup>b</sup>	7,328	1.05	1.00	0.99 <sup>b</sup>
CHL	3,457	1.37			13	4.61		
DEU	2,983	1.07	1.00	0.97 <sup>c</sup>	12,747	1.08	1.00	1.03 <sup>c</sup>
GBR	295	1.06	1.38	1.99 <sup>c</sup>	11,854	0.95	1.00	0.96 <sup>c</sup>
BEL	213	1.06	0.51	1.67 <sup>c</sup>	2,765	1.06	0.46	0.96 <sup>c</sup>
NLD	203	1.08	1.28	1.48 <sup>c</sup>	3,297	1.04	0.80	1.18 <sup>c</sup>
CHE	42	0.64	1.00		1,623	1.08	1.00	
CAN	14	1.11	4.38		2,973	1.02	1.33	
JPN	4	1.19	1.39		1,464	1.13	1.82	
MAD** [%]		18.0	44.0	32.3		70.6	27.3	75.0

+ 635 %

- 54 %

a) Australian Bureau of Statistics b) USDA FAS c) Eurostat - ComExt

\*) UN C = UN Comtrade \*\*) MAD = Mean absolute deviation

SITC-code 11217 (2006)

# Results Test II: Summary

## alternative data sources

---

- range of deviations:  
SITC 1121: -45% (CHE, OECD) to +293% (CAN, OECD)  
SITC 11217: -54% (BEL, OECD) to +635% (ESP, Eurostat)
- OIV data show smallest deviation from UN Comtrade export data, regardless of SITC-code
- extreme relative deviations seem to be more likely for small reported trade volumes
- Which data source can we trust?
- What are the reasons for differences although the same classification is used?

# Test III: Discrepancies between bilateral trade data (SITC 1121, 2000 – 2008)

Morgenstern (1950):

Case 1: country  $i$  ( $j$ ) overstates (understates) its foreign trade statistics

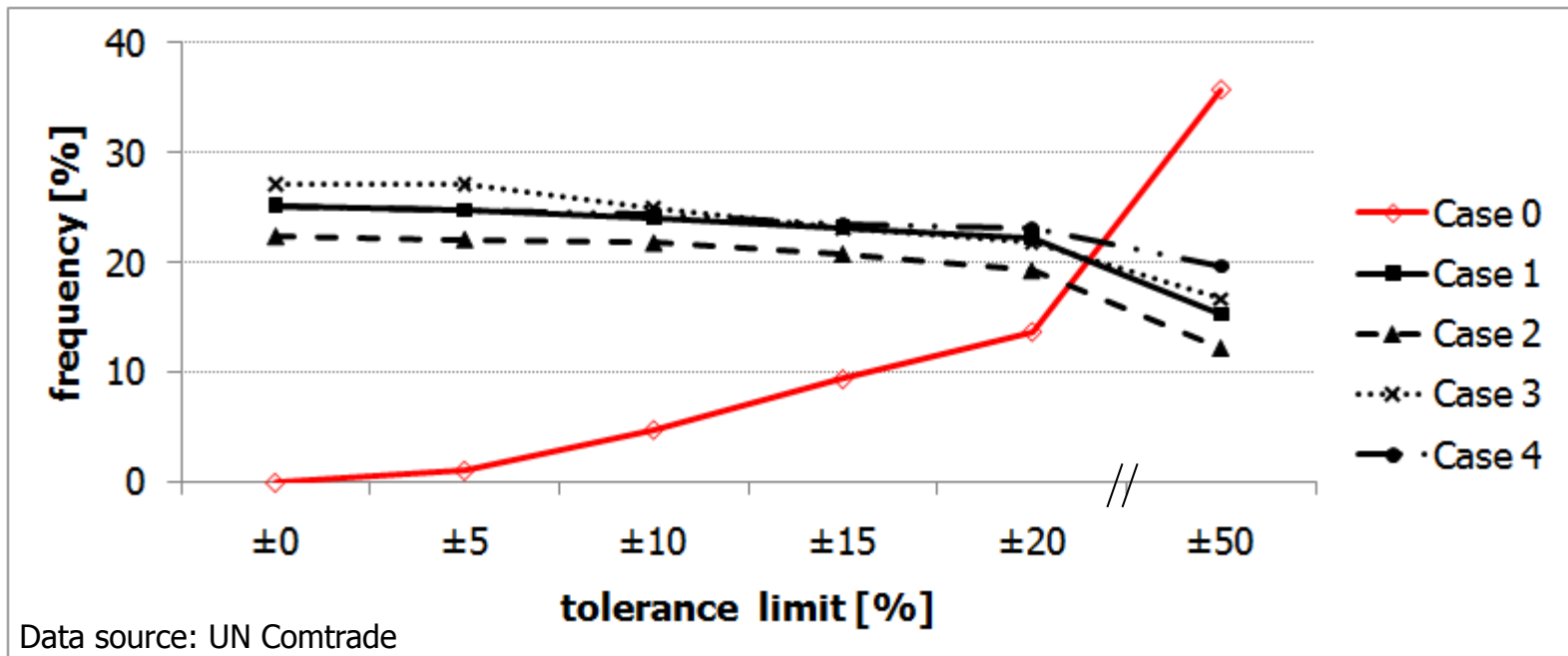
Case 2: imports overstated or exports understated in both countries

Case 3: exports overstated or imports understated in both countries

Case 4: country  $j$  ( $i$ ) overstates (understates) its foreign trade statistics

*Case 0: allow an accepted tolerance level*

## Distribution of cases 0 to 4 (SITC 1121, 2000-2008, n=1232)



# Results Test III: Summary

## bilateral trade data

---

- Trade data are often not identically reported by exporters and importers
- Case 1 to 4 nearly equally distributed if limit of tolerance is set to 0
- Nearly 65% of the bilateral trade data deviate from each other by more than  $\pm 50\%$
- Only for a few country-combinations a degenerate distribution of cases 1 to 4 exists  
(e.g. AUS-FRA, AUS-GBR, AUS-DEU, BEL-NLD, CHE-USA, CHE-GBR, CHL-DEU, CHL-FRA, CHL-USA)
- No structure of derivations detected

# Conclusion

---

- Discrepancies in trade data exist
  - between different data sources
  - within single data source
- No reason to prefer one data source over others
- No clear systematic bias detected
- Incorrect data used in studies may lead to misspecified models, wrong results or incorrect political implications
- **Take care while using wine trade data!!!**



—these figures I have given are wrong. I have given them because we do not have the right figures—

*wine is not unique with respect to data imperfections*

Franziska Thiemann, M.Sc.

Department of Agricultural Economics  
Section Innovation and Information  
Christian-Albrechts-University at Kiel  
Wilhelm-Seelig-Platz 6/7  
24098 Kiel, Germany

Email: [fthiema@ae.uni-kiel.de](mailto:fthiema@ae.uni-kiel.de)

Tel.: + 49 (0) 431 880 4412

Fax: + 49 (0) 431 880 4550

# Backup - Results Test II: Comparison of alternative data sources

Country	Export				Import			
	UN Comtrade [1,000 hl]	Ratio			UN Comtrade [1,000 hl]	Ratio		
		OIV / UN C*	OECD / UN C*	local Statistics / UN C*		OIV / UN C*	OECD / UN C*	local Statistics / UN C*
ITA	19,477	0.94	1.04	0.86 <sup>c</sup>	1,500	0.98	1.05	1.12 <sup>c</sup>
FRA	14,795	0.99	1.58	0.95 <sup>c</sup>	5,532	0.96	0.55	0.95 <sup>c</sup>
ESP	8,461	1.69	1.00	1.59 <sup>c</sup>	251	1.59	1.00	2.22 <sup>c</sup>
AUS	7,626	1.00	0.94	0.99 <sup>a</sup>	312	0.96	2.01	0.99 <sup>a</sup>
USA	3,839	0.98	1.00	0.98 <sup>b</sup>	7,911	0.98	1.00	0.99 <sup>b</sup>
CHL	3,514	1.35			17	3.65		
DEU	3,486	0.92	1.00	0.92 <sup>c</sup>	14,438	0.95	1.00	0.98 <sup>c</sup>
GBR	316	0.99	1.44	1.94 <sup>c</sup>	13,003	0.87	0.99	0.94 <sup>c</sup>
BEL	250	0.90	0.72	1.53 <sup>c</sup>	3,210	0.92	0.61	0.91 <sup>c</sup>
NLD	243	0.90	1.52	1.35 <sup>c</sup>	3,477	0.98	0.84	1.15 <sup>c</sup>
CHE	43	0.62	1.00		1,808	0.97	1.00	
CAN	16	0.96	3.93		3,078	0.99	1.35	
JPN	5	1.06	1.28		1,753	0.95	2.00	
MAD** [%]		14.7	42.7	30.0		28.8	28.5	19.3

a) Australian Bureau of Statistics b) USDA FAS c) Eurostat - ComExt

\*) UN C = UN Comtrade \*\*) MAD = Mean absolute deviation

SITC-code 1121 (2006)

# Backup – Results Test III

## Discrepancies between bilateral trade data (SITC 1121, 2000 – 2008)

Country pair	cases					HI	MADU-0-1
	0	1	2	3	4		
AUS-BEL				4	2	0.6	0.7
AUS-CAN	1	4	2			0.6	0.7
AUS-CHE		6	3			0.6	0.7
AUS-CHL		1	1	2	3	0.3	0.3
AUS-DEU		9				1.0	1.0
AUS-ESP	2	1	5		1	0.6	0.6
AUS-FRA			9			1.0	1.0
AUS-GBR		9				1.0	1.0
AUS-ITA			7		2	0.7	0.7
AUS-JPN			3			1.0	1.0
AUS-NLD		2	1	3	2	0.3	0.2
AUS-USA		8	1			0.8	0.9
$\Sigma$ (N=94)	3	40	32	9	10		
%	3.2	42.6	34.0	9.6	10.6		

Country pair	cases					HI	MADU-0-1
	0	1	2	3	4		
CHE-AUS			3		6	0.6	0.7
CHE-BEL			3		6	0.6	0.7
CHE-CAN		3	5		1	0.4	0.5
CHE-CHL			3			1.0	1.0
CHE-DEU				4	5	0.5	0.7
CHE-ESP			2		7	0.7	0.7
CHE-FRA				8	1	0.8	0.9
CHE-GBR				9		1.0	1.0
CHE-ITA		1	4		4	0.4	0.5
CHE-JPN				1	4	0.7	0.7
CHE-NLD				1	8	0.8	0.9
CHE-USA					9	1.0	1.0
$\Sigma$ (N=98)	0	4	20	23	51		
%	0.0	4.1	20.4	23.5	52.0		



# Backup – Morgensterns` discrepancy classes

---

$$Q_1 = \frac{M_{ji} - X_{ij}}{M_{ji}} * 100 \qquad Q_2 = \frac{X_{ji} - M_{ij}}{X_{ji}} * 100$$

$X_{ij}$  exports from country  $i$  to country  $j$  reported by country  $i$   
 $M_{ji}$  imports by country  $j$  from country  $i$  reported by country  $j$

Case 1:  $Q_1 > 0$  and  $Q_2 > 0$ . In this case either country  $i$  overstates its foreign trade statistics for exports and for imports, or country  $j$  understates its foreign trade in both for exports and imports.

Case 2:  $Q_1 > 0$  and  $Q_2 < 0$ . This may indicate that imports are overstated in both countries or that exports are understated in both countries.

Case 3:  $Q_1 < 0$  and  $Q_2 > 0$ . This may indicate that exports are overstated in both countries or that imports are understated in both countries.

Case 4:  $Q_1 < 0$  and  $Q_2 < 0$ . This may indicate that the second country overstates its foreign trade statistics in both directions or that the first country understates its foreign trade statistics in both directions.