



# Datenverknüpfung und Modellierung im Bereich der makroökonomischen Statistik

Nutzung von Verknüpfungspotenzialen für die Schätzung des  
Umsatzes in der Statistik der Unternehmensgruppen (STAGRE)

**Marius Ley, Bundesamt für Statistik**

Schweizer Tage der öffentlichen Statistik, 26. Oktober 2022



## Inhalt

1. Kontext: Statistik der Unternehmensgruppen und Umsatz
2. Potentielle statistische Quellen für den Umsatz
3. Lösungsansatz: Modellgestützte Schätzungen
4. Bestimmung des am besten geeigneten Modells
5. Favorisiertes Modell: Spezifikation
6. Resultate der Modellevaluation
7. Diffusion, Ausblick und Schlussfolgerungen



## Kontext: Statistik der Unternehmensgruppen

- Population:** Unternehmensgruppen in der Schweiz (Multinationale, aber auch rein inländische). Kriterium: Kontrollverhältnisse zwischen Einheiten
- Inhalte:** strukturelle (Beschäftigung, Anzahl Unternehmen und Arbeitsstätten) und monetäre Variablen
- Entstehung:** erstmals 2017 veröffentlicht, seither schrittweise erweitert
- Quellen (für Kontrollverhältnisse):** externe Datenlieferanten (inkl. EuroGroups Register EGR), interne Quellen (BUR), Administrativdaten (Mehrwertsteuer- und AHV-Gruppen), öffentlich verfügbare Quellen (Handelsregister, Medienmitteilungen, Jahresberichte)
- Publikation:** jährlich, jeweils im November (aktuellste Resultate für das Jahr T-1)



## Kontext: Statistik der Unternehmensgruppen

### Einige Eckwerte (2019):

- 52 819 Unternehmen  
(beinhalten 4,8% aller Unternehmen mit Beschäftigung gemäss STATENT)
- 1 925 462 Beschäftigte  
(mehr als einem Drittel der Beschäftigung gemäss STATENT)
- 239 033 Mio. Franken Warenimporte und  
286 224 Mio. Franken Warenexporte  
(mehr als 85% bzw. 90% des gesamten Handelsvolumens)
- 2 237 694 Mio. Franken Umsatz



## Kontext: Umsatz in der STAGRE

### Ziele der Bereitstellung des Umsatzes:

- Datengrundlage sowohl für strukturelle als auch für makroökonomische (gemäss dem Regelwerk der VGR) Analysen
  - Bedingt ausreichend hohen Detaillierungsgrad; z.B. nach Branchen, Art der Unternehmensgruppe und weiteren Aspekten
- Vergleichbarkeit mit anderen Ländern (insb. EU); konzeptuelle Basis: EU-Verordnungen zu SBS (250/2009) bzw. zu EBS (2019/2152)

### Herausforderungen:

- Keine unmittelbar verwendbare statistische Quelle für Umsatz (im Gegensatz zu Beschäftigung)
- Zusätzliche Erhebungen stellen keine Option dar

Gleiche Herausforderungen für eine Reihe von weiteren monetären Variablen



# Potentielle statistische Quellen für den Umsatz

Quellen	Abdeckung	WS erfasst nur 11% der STAGRE-Einheiten (punkto Besch. ca. 2/3)	Übereinstimmung mit EBS-Definition
Wert-schöpfungs-statistik ( <b>WS</b> )	<p><u>Universum</u>: <b>gut</b> (bestimmte Kombinationen von NOGA und Rechtsform sowie Einheiten mit &lt; 3 Beschäftigten ausgeschlossen)</p> <p><u>Nettostichprobe</u>: <b>eingeschränkt</b></p>		<p><b>Gut:</b></p> <ul style="list-style-type: none"> <li>- Unterscheidung zwischen <b>T010</b> (Umsatz betriebswirtschaftlich) und T011 (Übrige betriebliche Erträge) <b>im Sinne der SBS</b></li> </ul>
Mehrwert-steuer ( <b>MWST</b> )	<p><u>Universum</u>: <b>gut</b></p> <p><u>Effektive Abdeckung</u>: <b>bedingt</b> (von der MWST ausgenommene Leistungen; Einheiten mit &lt; CHF 100'000)</p>	<p><b>MWST-Umsatz für 2/3 der STAGRE-Einheiten (punkto Besch. &gt; 98%)</b></p>	<p><b>Mangelhaft:</b></p> <ul style="list-style-type: none"> <li>- Übrige betriebliche Erträge grössten-teils in MWST-Umsatz enthalten</li> <li>- Exporte teilweise nicht deklariert</li> <li>- Problematisch: «Verteilte» Umsätze (von MWST-Gruppen)</li> </ul>



## Lösungsansatz: Modellgestützte Schätzungen

Aggregation, innerhalb der gewünschten Gliederungsstufen (z.B. NOGA-Abteilungen), von

- **erhobenem** Umsatz der WS, falls vorhanden; und
- **Schätzwert** für den Umsatz, falls nicht vorhanden, basierend auf Modell mit den folgenden potentiellen Hilfsvariablen:
  - MWST-Umsatz
  - Beschäftigung gemäss AHV
  - Arbeitseinkommen gemäss AHV

Für die Schätzung und Evaluation des Modells werden **alle** Beobachtungen der WS (N = 14'439, Referenzjahr 2019) berücksichtigt, nicht nur jene des STAGRE-Universums (N = 5'838)  
➤ Höhere Effizienz der Schätzung



## Bestimmung des am besten geeigneten Modells – Anforderungen

- 1) minimale Abweichung zwischen Schätzwerten und **Originalwerten** der Nettostichprobe  
(auf Ebene der einzelnen Beobachtungen)
- 2) minimale Abweichung zu Resultaten der **direkten Schätzungen** der WS  
(auf Ebene NOGA-Abteilung\*; nach Einsetzung der erhobenen Werte, wo vorhanden)
- 3) minimale unplausible (modellbedingte) Ausschläge **im Zeitverlauf**  
(auf Ebene NOGA-Abteilung\*; nach Einsetzung der erhobenen Werte, wo vorhanden)





# Bestimmung des am besten geeigneten Modells – Implementierung

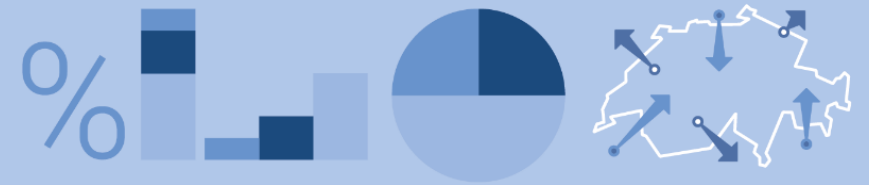
Qualitätsmasse:

1) **QM-Global**: Präzision eines Modells auf Ebene der einzelnen Beobachtungen (**Anforderung 1**)

$$RMSE_w = \sqrt{\frac{1}{N} \sum_{i=1}^N [(\hat{y}_i - y_i^R) w_i]^2}$$

2) **QM-Aggregat**: Präzision eines Modells auf Ebene der Resultate auf Ebene der NOGA-Abteilungen\*. z-Werte; konkret Mittelwert bzw. mittleres Quadrat daraus (**Anforderung 2**)

**Graphische Darstellung** im Zeitverlauf und Beurteilung (**Anforderung 3**)



## Favorisiertes Modell: Spezifikation

Unterstellte Funktion für  $U_i$  (Umsatz gemäss WS) jeder Einheit  $i$ :

$$U_i = A_{g(i)} T_i^{\beta_{g(i)}} L_i^{\gamma_{g(i)}} e^{\varepsilon_i}$$

- $T_i$ : **Umsatz** gemäss MWST
- $L_i$ : **Arbeitseinkommen** gemäss AHV/STATENT
- $g(i)$ : Zugehörigkeit zu einer für die Modellierung relevanten **Subgruppe** (insbesondere: NOGA-Abteilungen, NOGA-Arten)

**Intuition:** WS-Umsatz verhält sich proportional zu (multiplikativer) Kombination aus MWST-Umsatz und Arbeitseinkommen, wobei die relativen Gewichte dieser beiden Variablen ( $\beta$  und  $\gamma$ ) und der Skalierungsfaktor ( $A$ ) zwischen den Subgruppen (Branchen etc.) variieren können.



## Favorisiertes Modell: Spezifikation

Unterstellte Funktion für  $U_i$  (Umsatz gemäss WS) jeder Einheit  $i$ :

$$U_i = A_{g(i)} T_i^{\beta_{g(i)}} L_i^{\gamma_{g(i)}} e^{\varepsilon_i}$$

- «**Mixed Model**»: Random-Koeffizienten ( $\beta_{g(i)}$ ,  $\gamma_{g(i)}$ ) und Random-Effekte (in  $A_{g(i)}$ ) sind hierarchisch spezifiziert (NOGA-Abteilung / NOGA-Art).
- $A_{g(i)}$  beinhaltet einige **Fixeffekte**.
- Schätzung erfolgt linear (**log-log**; erfordert Skalierung für Rücktransformation der Schätzwerte).
- Vereinfachtes Modell (ohne  $T_i$ ) für Schätzung des Umsatzes von Einheiten **ohne MWST-Umsätze**



## Resultate der Modellevaluation: Qualitätsmasse

Modellvariante	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Fix- statt Random-Effekte; nur für NOGA-Abteilung	Wie (1), aber VZÄ gemäss STATENT (statt Arbeits-einkommen) als zweite Hilfsvariable	Random-Effekte nur für NOGA-Abteilung	<b>Favorisiertes Modell (Basis)</b>	Wie (4), aber Verzicht auf Arbeits-einkommen als Hilfsvariable (wo MWST-Umsatz vorhanden)	Wie (4), aber unter Verwendung einer weniger konservativen Skalierung	Wie (4), mit einigen Verfeinerungen und punktuellen Korrekturen von Einheiten
<b>QM-Global</b> (gewichteter RMSE zwischen modellbasierten Schätzungen und erhobenen Werten, in CHF Mio, N = 99'187)							
	35.50	36.14	35.01	31.21	33.35	31.36	30.82
<b>QM-Aggregat: Mittelwert</b> (der z-Werte der auf Ebene noga2r aggregierten Schätzungen, N = 419)							
Mittelwert	-0.166	-0.163	-0.099	-0.039	0.805	0.050	-0.132
p-Wert	0.001	0.001	0.043	0.432	0.000	0.307	0.007
<b>QM-Aggregat: RMS</b> (der z-Werte der auf Ebene noga2r aggregierten Schätzungen, N = 419)							
RMS	1.923	1.973	2.012	1.987	3.134	2.041	1.739
p-Wert (chi <sup>2</sup> )	0.000	0.000	0.000	0.000	0.000	0.000	0.000



## Resultate der Modellevaluation: Qualitätsmasse

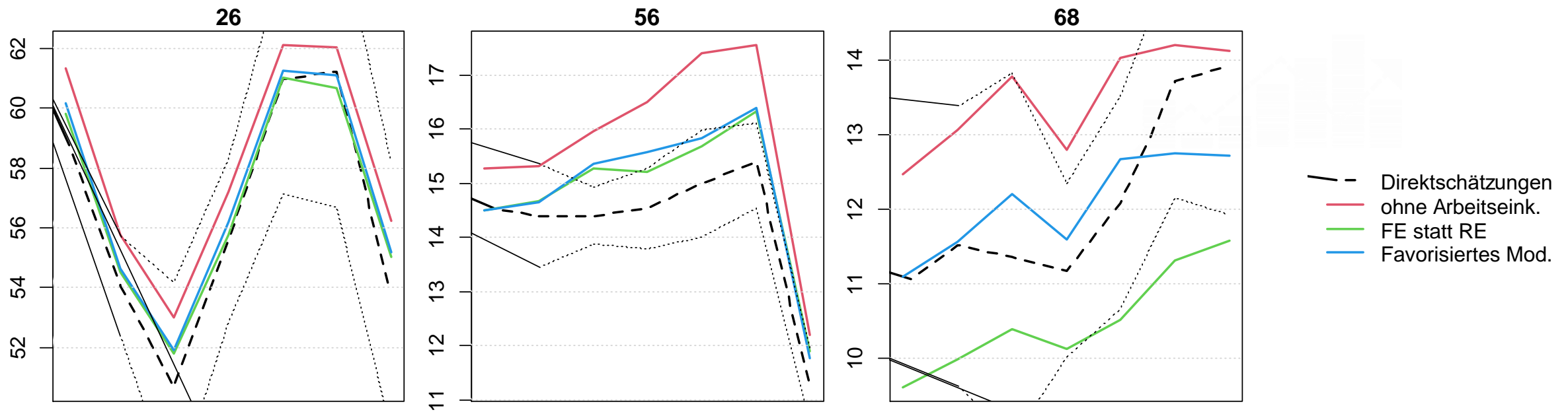
Modellvariante	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Fix- statt Random-Effekte; nur für NOGA-Abteilung	Wie (1), aber VZÄ gemäss STATENT (statt Arbeits-einkommen) als zweite Hilfsvariable	Random-Effekte nur für NOGA-Abteilung	<b>Favorisiertes Modell (Basis)</b>	Wie (4), aber Verzicht auf Arbeits-einkommen als Hilfsvariable (wo MWST-Umsatz vorhanden)	Wie (4), aber unter Verwendung einer weniger konservativen Skalierung	Wie (4), mit einigen Verfeinerungen und punktuellen Korrekturen von Einheiten
<b>QM-Global</b> (gewichteter RMSE zwischen modellbasierten Schätzungen und erhobenen Werten, in CHF Mio, N = 99'187)							
2014	37.509	37.799	37.452	28.333	31.912	28.434	28.098
2015	28.688	29.793	28.619	26.145	29.186	26.385	26.119
2016	35.986	36.248	35.887	30.887	33.991	31.274	30.519
2017	32.053	32.814	32.388	28.801	30.229	28.847	28.334
2018	36.432	36.796	36.621	34.742	35.588	34.935	34.005
2019	37.055	38.285	38.422	35.709	38.354	35.781	35.143
2020	38.971	39.633	34.027	32.100	33.031	32.159	31.911



# Resultate der Modellevaluation: graphisch, 2014–2020

In Mio. Franken. Auswahl einiger Branchen:

26 (Datenverarbeitungsgeräte und Uhren), 56 (Gastronomie), 68 (Immobilienwesen)





## Resultate der Modellevaluation

Weitere Alternativen wurden evaluiert (und verworfen), unter anderen:

- in Absolutwerten statt log-log
- unter Ausschluss von mutmasslichen Ausreissern
- mit zusätzlichen Fixeffekten
- mit weniger detaillierten und anders strukturierten Random-Effekten



## Resultate und Diffusion

**Beispiel:** Sitzländer (nur Top 7, nach Umsatz 2019) der Unternehmensgruppen

Sitzland	Umsatz, in Mio. Franken		Vertraulichkeitsregeln:			
	2014	2015	<ul style="list-style-type: none"> <li>• mindestens 4 Einheiten mit Umsatz; und</li> <li>• Einheit mit dem grössten Umsatz trägt weniger als 60% zum Zelltotal bei</li> </ul>			
<b>Total Unt.'gruppen</b>	<b>2 189 075</b>	<b>1 812 260</b>				
Schweiz	772 194	683 122				
Vereinigte Staaten	342 355	325 387	330 917	406 215	467 834	436 792
Russland	X	X	X	117 331	139 556	149 495
Deutschland	84 929	82 216	77 282	81 002	87 607	89 087
Liechtenstein	1 744	1 379	X	X	77 044	67 755
China	10 415	8 279	12 749	31 897	42 639	47 127
Vereinigtes Königreich	26 863	28 136	29 442	51 170	53 840	43 463





## Ausblick: Weitere Schritte

Im November 2022: Publikation ausführlicher Bericht (zeitgleich mit den neuesten Resultaten der STAGRE)

<https://www.bfs.admin.ch/bfs/de/home/statistiken/industrie-dienstleistungen/stagre.html>

Mittel- und langfristig:

- Produktion weiterer monetärer Variablen (insbesondere: Bruttoproduktionswert, Bruttowertschöpfung, Vorleistungen)
- Analysen zum Einbezug von Einheiten mit weniger als 3 Beschäftigten
- Eventuell Varianzschätzung
- Eventuell Produktion zeitnäherer Resultate (T-1 statt T-2)



## Schlussfolgerungen

Bisherige Erfahrungen mit modellgestützten Schätzungen sind positiv:

- Nutzbarmachung der Daten einer existierenden Stichprobenerhebung (WS) sowie von Administrativdaten (MWST, AHV) für breitere statistische Anwendung
- Die hier verwendeten Quellen könnten – wenn isoliert voneinander verwendet – die weiter oben definierten Ziele nicht erfüllen.
- Datenverknüpfungen ermöglichen signifikanten statistischen Mehrwert.

Wesentliche Herausforderungen: Modellspezifikation, -evaluation und Dokumentation sind aufwändig.



# Besten Dank für das Interesse!

## Fragen, Anregungen?

