



Jobinserate als Daten- quelle zur Identifikation von Fachkräftemangel

Claudia Plaimauer

3 Juli 2024

33F<



Jobinserate als Datenquelle zur Identifikation von Fachkräftemangel

Kontext

Fachkräftemonitoring in Österreich
BMAW AMS Fachkräftebarometer

Machbarkeitsstudie

Fragestellungen & Untersuchungsmethoden
Ergebnisse im Überblick

Online-Jobinserate

Big Data Plattform *Jobs Data*: Vorteile und Herausforderungen

Handlungsempfehlungen

Vorteilhafte Strategien / Grenzen der Interpretation

Kontext



Fachkräftemonitoring in Österreich

$$\text{Stellenandrang} = \frac{\text{Zahl der Arbeitslosen}}{\text{Zahl der offenen Stellen}}$$



Dargestellt in

- [BMAW Mangelberufsliste der Fachkräfteverordnung](#)
- [WKO Fachkräfteradar](#)
- Fachkräftemonitore von [OÖ](#) & [Tirol](#)
- [Wirtschaftsbund-Stellenmonitor](#)

BMAW AMS Fachkräftebarometer

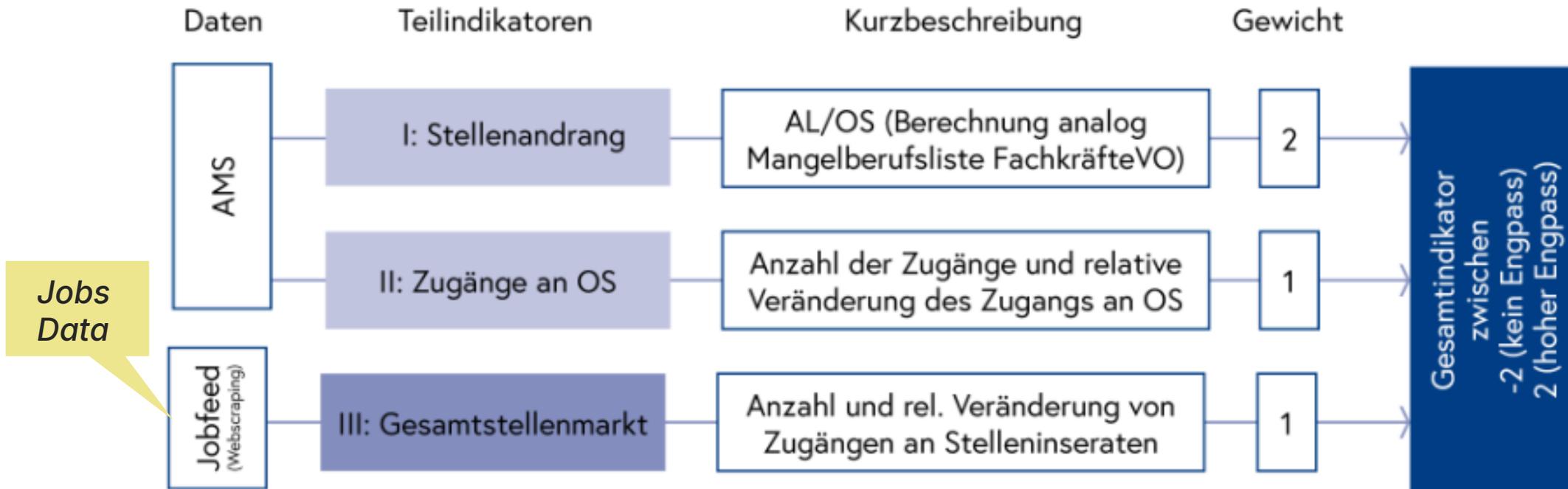


- Mai 2023: Veröffentlichung eines neuen Engpassindikatoren-Modells unter <https://www.bmaw.gv.at/Themen/Arbeitsmarkt/Arbeitsmarktdaten/Fachkraeftebarometer.html>
- Zeigt zeitnahe quartalsweise Ergebnisse zu Fachkräftenengpässen auf Berufsebene für Österreich und die einzelnen Bundesländer.
- Bildet als Sammelindikator mit drei Teilindikatoren sowohl kurzfristige saisonale sowie strukturelle Niveauelemente ab, als auch konjunkturelle Entwicklungen.

BMAW AMS Fachkräftebarometer

Fachkräftebarometer auf Basis von 3 Teilindikatoren

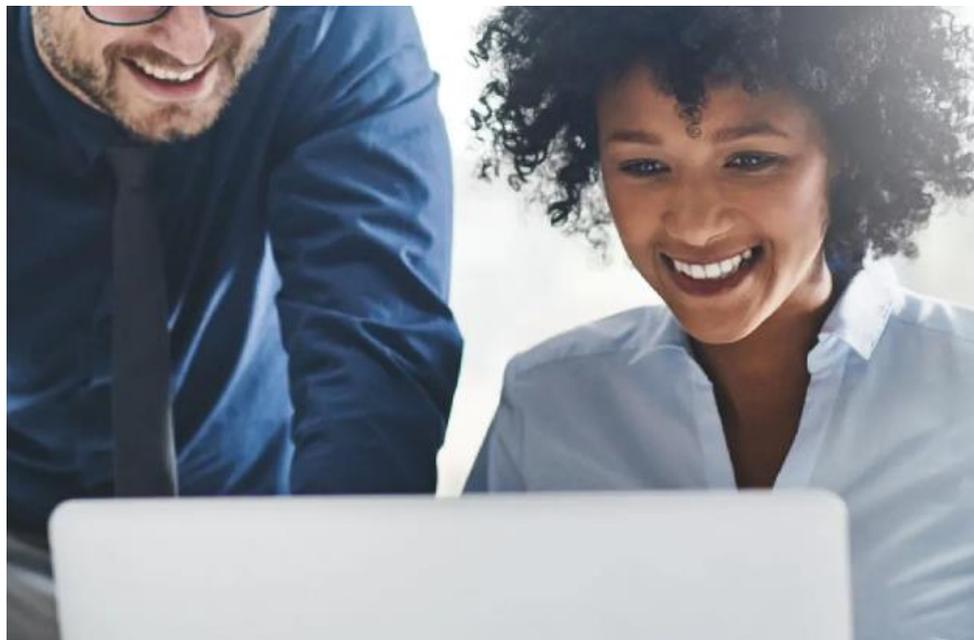
OS = offene Stellen
AL = Arbeitslose



Online-Jobinserate



Big Data Plattform *Jobs Data*



- Existiert seit 2003 in NL; nach und nach auch für 14 weitere nationale Jobmärkte umgesetzt (für Ö seit 03/2015)
- Laufende Verbesserung von Marktabdeckung und Datenqualität
- AMS Österreich ist Schlüsselkunde, daher zusätzliche Abfragemöglichkeiten mittels AMS-spezifischer Klassifikationen für Berufe, Kompetenzen, Ausbildungsniveaus und Regionen
- Abdeckung des österr. Online-Stellenmarktes liegt derzeit bei geschätzten 95-97%.

<https://www.textkernel.com/products-solutions/labour-market-insights/jobs-data/>

Vorteile

- Unabhängigkeit von Primärerhebungen
- (annähernde) Vollerhebung
- Zeitnahe Verfügbarkeit
- Detailreichtum (Jobtitel, Anforderungen, Arbeitgeber, Gehaltsangaben u.v.m.)
- Flexible Abfrageoptionen (Freitext, systematisch, beliebige Zeitfenster)
- Zugänglich über Portal, API, Datenfeed
- Hohe Transparenz

Herausforderungen

- Vollständigkeit und Genauigkeit der automatisiert gesammelten, extrahierten und normalisierten Jobinserate?
- Wie von der Anzahl beobachteter Jobinserate auf die nicht beobachtete Anzahl offener Stellen schließen?
- Wie zuverlässige Zeitreihen erstellen?
- Wie Repräsentativität gewährleisten?

Machbarkeitsstudie



Fragestellungen

- Wie plausibel sind die in der Pilotanalyse ermittelten Ergebnisse?
- Wie zuverlässig ist in *Jobs Data* die automatisierte Normalisierung der Jobinserate auf Bundesländer sowie ISCO-Berufsgattungen?
- Mit welcher Methode könnte man von der Anzahl veröffentlichter Jobinserate auf die Menge tatsächlich offener Stellen schließen?
- Wie könnte man Zeitverläufe darstellen, die nicht durch diverse technische (z.B. verbesserte Crawler, Serverausfälle) oder organisatorische Einflussfaktoren (z.B. verändertes Inserierungsverhalten) verzerrt werden?
- Welche Berechnungsmethode (Zufluss, Bestand) ist zuverlässiger?



Methoden

Datenvergleiche:

- *Jobs Data*-Ergebnisse aus der Pilotauswertung (Bestand 2019-2021; Zugang 2019-2023/Q1), zusätzliche Auswertungen nach Einzelmonaten & Quartalen (Durchschnitt mehrerer Stichtage)
- Offene-Stellen-Erhebung, AMS-Stellendaten, Österreich-Ergebnissen von Skills OVATE (Cedefop), Wirtschaftsbund-Stellenmonitor

Stichproben-Überprüfung der in *Jobs Data* vorliegenden Normalisierungen

Literaturstudie & ExpertInnen-Befragung



Jobs Data (Bestand) im Vergleich zu anderen Quellen

- Ursprüngliche Berechnungsmethode: Akkumulierende Zählung sämtlicher innerhalb eines Quartals als „aktiv“ geführter Inserate
- Verzerrung durch Re-Postings und (fälschlich?) noch als „aktiv“ geführte Inserate älteren Datums
- Vergleichbarkeit mit alternativen Datenquellen meist nur möglich hins. prozentueller Verteilungen und Verläufe, nicht hins. Quantitäten (unterschiedliche Zähleinheiten und Marktabdeckung)

Ergebnisse

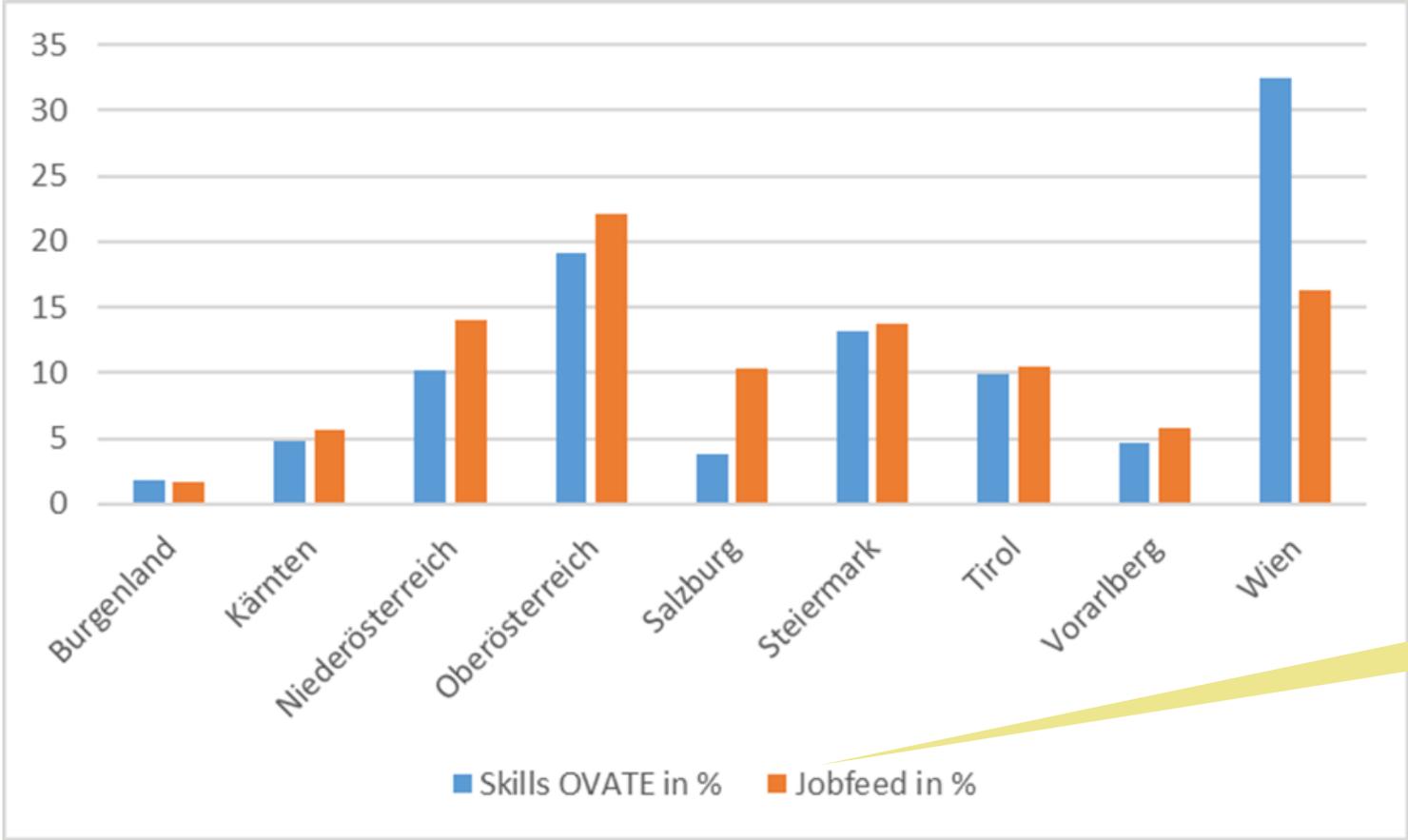
- Bessere Berechnungsmethode: Durchschnitt einzelner Stichtags-Bestände
- Hins. Marktabdeckung ist *Jobs Data* dem Wirtschaftsbund-Stellenmonitor und Skills OVATE überlegen
- Historische Daten des AMS eJob-Room (\neq AMS-Stellenstatistik) derzeit nur in *Jobs Data* systematisch abfragbar
- Das Jobangebot zusätzlich zu den AMS-Stellen wurde bisher unterschätzt

Bestand offener Inserate-/Stellen 2023/08

Erhebung	AMS-Stellen	AMS eJob-Room-Inserate	Weitere Stellen	Weitere Inserate	Alle Stellen	Alle Inserate
AMS-Stellenstatistik	109.813	-	-	-	-	-
Jobs Data	-	119.391	-	125.310	-	244.701
Offene-Stellen-Erhebung*	120.500	-	81.800	-	202.300	-
Wirtschaftsbund-Stellenmonitor	109.813	-	-	105.139	214.952	214.952

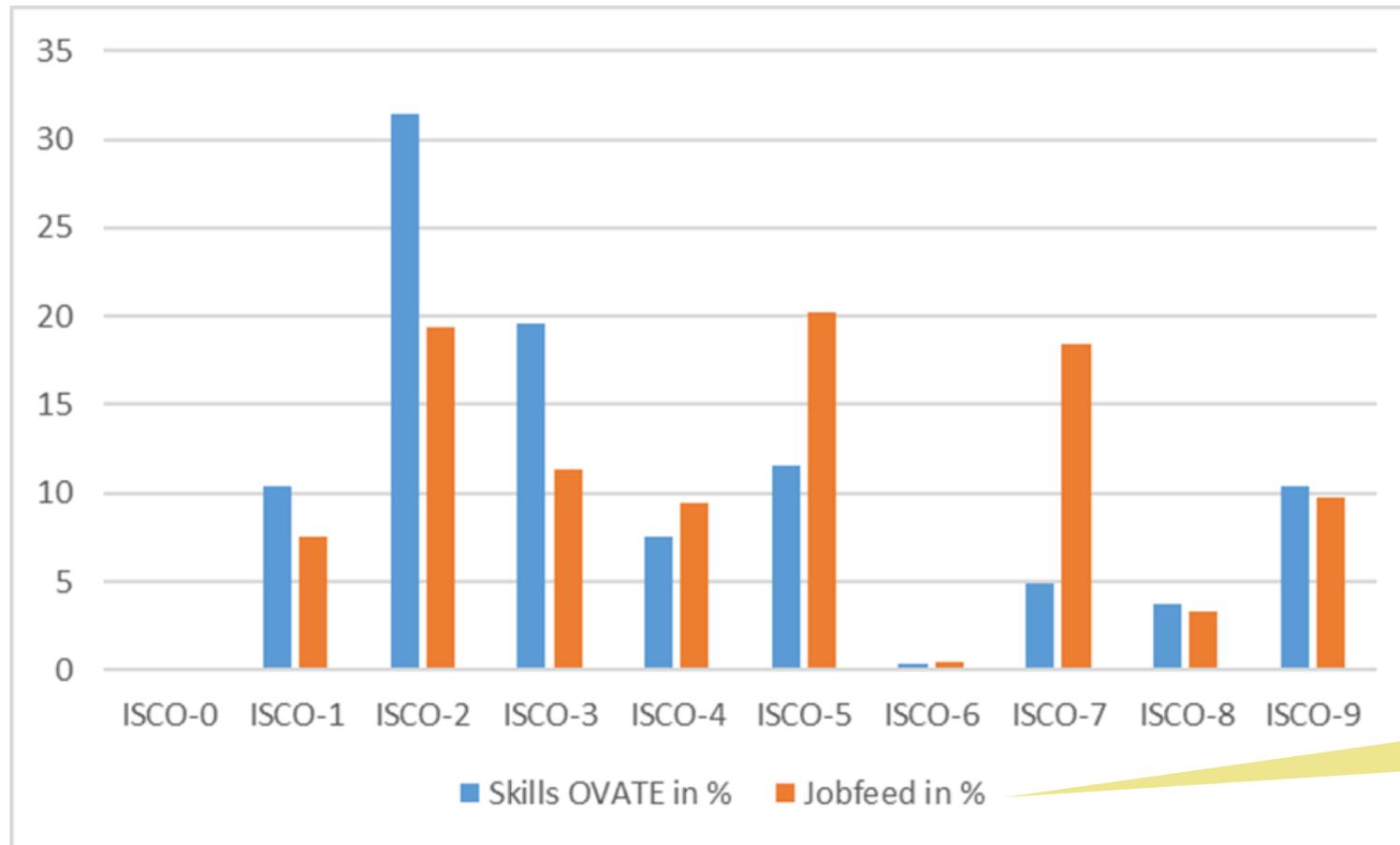
* Ergebnisse für 2023/Q3

Bestand offener Inserate 2021/Q4 – Bundesländer-Anteil



Jobs
Data

Bestand offener Inserate 2021/Q4 – Anteil ISCO-1-Steller



*Jobs
Data*

Bestand offener Inserate-/Stellen mit AMS-Beteiligung: Bundesländer-Anteil

Bundesland	Bestand AMS eJob-Room Inserate in % lt. <i>Jobs Data</i> (Stichtag 15.3.2023)	Bestand offene Stellen in % lt. AMS-Stellenstatistik in %
Burgenland	2	1
Kärnten	8	7
Niederösterreich	16	17
Oberösterreich	26	25
Salzburg	9	9
Steiermark	15	14
Tirol	9	7
Vorarlberg	4	4
Wien	12	14

Qualität der BL-Normalisierung

- In welchem Ausmaß kommt eine redaktionell vorgenommene Regionalisierung der Inserate zu einem vergleichbaren Ergebnis wie die automatisierte?
- Zufällige Stichprobe von 500 Inseraten (ca. 0,014% der Datenbasis)
- Von den 500 Inseraten der Stichprobe konnten 402 automatisiert einem Bundesland zugewiesen werden, die restlichen 98 Inserate nicht. („Unbekannt/Ausland“)

Ergebnisse

- *Jobs Data*-Inserate, die automatisiert einem Bundesland zugeordnet worden sind, sind zu einem hohen Prozentsatz (92% der Stichprobe) passend regionalisiert.
- Die Bundesländer-Normalisierung ist allerdings nicht ganz vollständig, denn Inserate, die automatisiert gar nicht bzw. auf „Unbekannt/Ausland“ normalisiert worden waren, konnten redaktionell für 30,6% der Stichprobe sehr wohl einem Bundesland zugeordnet werden.

Qualität der ISCO-Normalisierung

- Auffällige Ergebnisse in den *Jobs Data*-Bestandsdaten für 2021 gesucht: ISCO-Berufsgattungen ohne bzw. mit zu hohem oder zu niedrigem Inserateaufkommen.
- Redaktionelle Überprüfung auffälliger ISCO-Berufsgattungen

Ergebnisse

- Der Jobtitel der Inserate ist oft zu vage oder zu mehrdeutig, um auf dieser Basis eine eindeutige Zuordnung zu einem ISCO-Viersteller vorzunehmen (gilt für die automatisierte Normalisierung wie für die redaktionelle)
- V.a. eine trennscharfe Unterscheidung von ISCO-Vierstellern, die sich nur hinsichtlich Skill-Level unterscheiden, ist kaum möglich, daher höchste Vorsicht bei Aussagen auf Ebene von ISCO-Berufsgattungen!
- Wenn laut *Jobs Data* allerdings für manche ISCO-Viersteller kein Inserateaufkommen ausgewiesen wird, so kann diesem Ergebnis zum größten Teil vertraut werden.

Inserate → Stellen?

- Methodendiskussion auf europäischer Ebene: Maximum likelihood estimation, um von Online-Inseraten auf Stellen hochzurechnen (Utopie?)
- Spannungsindikator des UWV (NL): Nachträgliches Gewichten und Skalieren von annähernd flächendeckend erhobenen Online-Inserate-Daten, um Repräsentativität herzustellen und um auf die Anzahl tatsächlich offener Stellen zu schließen
- Stellenmarkt-Monitor der Universität Zürich (CH): Sammeln und Auswerten einer repräsentativen Stichprobe von Online-Inseraten; Anzahl offener Stellen bzw. gesuchter Personen wird reaktionell ermittelt

Ergebnisse

Will man auf Basis von *Jobs Data* von Inseraten auf Stellen schließen, wären dafür größere Investitionen nötig, etwa in

- Entwicklung und Testung statistischer Methoden, um Online-Inserate auf Stellen hochzurechnen.
- Eine Betriebsbefragung, die auch auf Detailebene repräsentativ ist (→Basis eines Gewichtungsmodells).
- Eine Erhebung, die eine repräsentative Stichprobe von Inseraten aus *Jobs Data* entnimmt und diese ergänzend codiert hinsichtlich Anzahl tatsächlich offerierter Stellen bzw. gesuchter Personen.

Zeitverläufe?

- *Jobs Data* wurde nicht vorrangig als Quelle für Bedarfsanalysen konstruiert (Repräsentativität & Datenkonsistenz haben keine Priorität)
- Verzerrungsfaktoren sind z.B. zunehmend bessere Marktabdeckung, zwischenzeitliche Unterbrechung von Datenflüssen, Weiterentwicklung von Extraktion und Normalisierung; Änderung betrieblicher Rekrutierungspraxis
- Lösungsansätze v.a. in Best-Practice-Beispielen und der aktuellen Methodendiskussion zu finden

Ergebnisse

Ansätze, um Datenkonsistenz über längere Zeitspannen herzustellen, sind z.B.

- Entwicklung eines Gewichtungs-Modells
- Ziehen einer repräsentative Zufallsstichprobe
- Optimieren von Datenerschließung und Qualitätsicherung für Forschungszwecke
- Genaues Monitoring der Datenstruktur
- Entwicklung statistischer Methoden, die wenig sensibel auf Datenausreißer reagieren und die Änderungen bei Webquellen oder in der Marktabdeckung glätten können

Handlungsempfehlungen



Vorteilhafte Strategien

- Analyse kurzer Zeitabschnitte (< 6 Wochen), weil sonst Verzerrung von Inseratebeständen durch Re-Postings.
- Qualität der Webquellen überprüfen, um Job-Portale zu entfernen, die kaum originäre Inserate enthalten sowie solche, die das Datum der Erstveröffentlichung verfälschen.
- Analyse des Inserate-Zugangs bevorzugen, weil in Bestandsanalysen eine Fehlinterpretation bereits abgelaufener Inserate als noch aktiv nicht ausgeschlossen werden kann.
- Anzahl ermittelter Jobinserate nicht in Bezug setzen zu Anzahl arbeitslos gemeldeter Menschen; stattdessen mit Vorjahreswerten vergleichen.
- Vorsicht mit Aussagen auf Ebene von ISCO-Berufsgattungen.



Danke für Ihre Aufmerksamkeit!



Claudia Plaimauer
3s Unternehmensberatung GmbH
Wiedner Hauptstraße 18
1040 Wien

T. +43-1-5850915-33
claudia.plaimauer@3s.co.at