

Dokumentation

Registerbasierte Erwerbsverläufe in der Covid-19 Pandemie (RECOV)

Version 1.0, Mai 2022

Inhalt

1	Allgemeines zur Datengrundlage	2
2	Grundgesamtheit und Stichprobe.....	3
3	Erwerbsverlaufs-Datensatz (recov-ewv_v1)	4
3.1	Demografische Variablen (Block A).....	4
3.2	Arbeitsmarktstatus im Zeitverlauf (Block B)	5
3.2.1	Erste Versicherungsperioden und Gesamtdauer ausgewählter Perioden	5
3.2.2	Monatlicher Arbeitsmarktstatus in der Pandemie 2019 bis 2021	5
3.3	Jährliche Statustage 2016 bis 2021 (Block C).....	6
3.4	Einkommen nach Einkommensarten 2016 bis 2021 (Block D)	6
4	Episodendatensatz (recov-eh_v1).....	8
5	Betriebsdatensatz (recov-dg_v1).....	8
5.1	Jährliche Indikatoren 2016 bis 2021	8
5.1.1	Lohnsumme und Durchschnittslohn pro Tag und Beschäftigte.....	8
5.2	Versicherungsträger, Standort und NACE-Sektor	9
5.3	Monatliche Betriebsindikatoren	9
6	Materialien und Ressourcen	9

1 Allgemeines zur Datengrundlage

Die Daten-Infrastruktur RECOV (Registerbasierte Erwerbsverläufe in der Covid-19 Pandemie) besteht aus drei miteinander verknüpfbaren Einzeldatensätzen.¹ Der Erwerbsverlaufs-Datensatz (recov-ewv_v1) besteht aus einer 5%-Zufallsstichprobe der österreichischen Wohn- und Erwerbsbevölkerung zu Beginn des Jahres 2020 (d.h. die in Österreich wohnhafte Bevölkerung zuzüglich den aus dem Ausland einpendelnden Erwerbspersonen) im Alter von 15 bis 64 Jahren (im Referenzjahr 2020).² Dieser Datensatz verfügt über einen Eintrag (Zeile) pro Person und stellt neben grundlegenden demografischen (Alter, Geschlecht, Mutterschaft, etc.) und regionalen (Region des Wohnortes, Staatsbürgerschaft, etc.) detaillierte Informationen über die Erwerbsverläufe (Arbeitsmarktstatus, Beschäftigung, Kurzarbeit, Arbeitslosigkeit, Einkommen, etc.) von 2016 bis 2021 bereit. Bei Beschäftigungseinträgen sind Informationen über die Betriebe (Wirtschaftsbereich und Standort) enthalten.

Der Episodendatensatz (recov-eh_v1) stellt für jede ausgewählte Person sämtliche sozialversicherungsrechtlichen Beschäftigungs-, Arbeitslosigkeits- und Nichterwerbsepisoden aggregiert nach Typen seit 2016 bereit. Die Anzahl der Einträge (Zeilen) pro Person entspricht der Anzahl der aufeinander folgenden Episodentypen, jeweils mit Beginn- und Enddatum. Der dritte Datensatz (recov-dg_v1) stellt zusätzliche Informationen über die Betriebe bereit: Lohnhöhe, Personalstand, Frauenanteil der Beschäftigten, Durchschnittsalter, Anzahl geringfügig Beschäftigter und Ausländeranteil).

Diese Datensätze wurden auf Basis verschiedener *Datentabellen* der Arbeitsmarktdatenbank (AMDB arbeitsmarktdatenbank.at) erstellt. Folgende AMDB-Datentabellen werden verwendet:

Daten der Sozialversicherung bzw. des Dachverbandes der österreichischen Sozialversicherungsträger (DV, vormals Hauptverband, HV)

1. Tabelle hv_epi_stamm_bas
2. Tabelle hv_pn
3. Tabelle hv_dg_konto
4. Tabelle hv_beitrg
5. Tabelle mon_mitvers_int

Verschneidung HV-Datenaufbereitung mit anderen Datenquellen

6. Tabelle mon_uni_status_int (Episodenaufbereitung, Quellen: DV, AMS, SMS)
7. Tabelle mon_ew_xt_uni_bus (Erwerbskarrierenmonitoring, Quellen: DV, AMS, SMS)
8. Tabelle mon_betrieb_bus (Betriebsmonitoring, Quellen: DV, AMS, SMS)
9. Tabelle mon_zusatz_bus (Zusätzliche Personenmerkmale, Quellen: DV, AMS, Synthesis)
10. Tabelle mon_mig_bus (Staatsangehörigkeit und Einbürgerung, Quellen: DV, AMS)

AMS-Zusatzdaten (aktuell nicht allgemein, sondern nur auf Projektbasis verfügbar)

11. Tabelle personen_bas

¹ Die AMDB-Personen- bzw. Betriebsnummern (PENR bzw. BENR) werden aufgrund deren Länge der Einfachheit halber jeweils durch Laufnummern ersetzt („lfpe“ bzw. „lfdg“). Die Zuordnungsschlüssel (PENR-lfpe bzw. BENR-lfdg) werden bereitgestellt.

² Es werden die Geburtsjahrgänge 1955 bis 2005 herangezogen, weil in den Daten nicht das Geburtsdatum, sondern das Geburtsjahr verzeichnet ist, weshalb das Alter zum Jahresbeginn nicht exakt bestimmt werden kann.

Die AMDB stellt grundsätzlich Informationen über sämtliche Versicherungszeiten bei einem dem Dachverband der österreichischen Sozialversicherungsträger (DV) angehörigen Sozialversicherungsträger und über die Vormerkzeiten beim Arbeitsmarktservice bereit.

Alle Beschäftigungsverhältnisse, die keine Pflichtversicherung bei einem DV-Träger auslösen sind damit **nicht bzw. nur teilweise erfasst**. Das betrifft Angehörige der **Kammern der freien Berufe**, Versicherte der **Krankenfürsorgeanstalten** sowie öffentlich Bedienstete des Bundes und der Bundesländer. Vollständig außerhalb des DV-Universums ist allerdings nur eine sehr kleine Minderheit der Erwerbstätigen: den rund 4 Mio. im DV versicherten Erwerbstätigen stehen rund 32 Tausend Erwerbstätige gegenüber, die ausschließlich außerhalb des DV erfasst sind (0,8%).³

Für weitere Informationen zu den einzelnen AMDB-Datentabellen (grundlegende Beschreibung, Variablenlisten, Datenquellen, verfügbarer Zeitraum, Dimensionstabellen, etc. der oben genannten Datentabellen 1 bis 10) siehe die offizielle Dokumentation (AMS - BMAFJ, 2020).⁴

2 Grundgesamtheit und Stichprobe

Grundgesamtheit ist die 15- bis 64-jährige Wohnbevölkerung in Österreich sowie die Arbeitskräfte mit ausländischem Wohnsitz, die in Österreich als Erwerbspersonen tätig sind. Zur Bestimmung der Grundgesamtheit werden die Jahrgänge der von 1955 bis 2005 Geborenen in den Tabellen „hv_pn“ und „mon_mitvers_int“ gesucht. Das ergibt in Summe 6.176.194 unterschiedliche Personen mit gültigem „hv_pn“-Eintrag am 15.2.2020 oder einer zu diesem Zeitpunkt aktiven Mitversicherung lt. „mon_mitvers_int“. Laut der Tabelle „mon_zusatz_bus“ ist der Wohnort von 206.375 Personen nicht in Österreich. Bleiben 5.969.819 Personen mit Wohnsitz in Österreich, was sehr nahe an der offiziellen Statistik des Bevölkerungsstandes ist. Denn laut Statistik des Bevölkerungsstandes lebten zu Jahresbeginn 2020 5.924.377 15- bis 64-Jährige in Österreich (siehe Tabelle 1).⁵

Tabelle 1: Vergleich Grundgesamtheit AMDB mit Statistik des Bevölkerungsstandes

Bevölkerungsstand (ZMR)		AMDB		
Alter	Hauptwohnsitz Ö	Geburtsjahre	Wohnort Ö	Inkl. Wohnort Ausland
15 bis 29 Jahre	1.557.027	1991 bis 2005	1,613,711	1,647,579
30 bis 44 Jahre	1.782.373	1976 bis 1990	1,768,762	1,840,470
45 bis 59 Jahre	2.019.342	1961 bis 1975	2,024,689	2,093,461
60 bis 64 Jahre	565.635	1955 bis 1960	566,298	598,858
15 bis 64 Jahre	5.924.377	1955 bis 2005	5,973,460	6,180,368

Quelle: Bevölkerungsstand (ZMR): Statistik des Bevölkerungsstandes (STATcube – Statistische Datenbank von STATISTIK AUSTRIA); AMDB: eigene Auswertungen auf Basis der Tabellen „hv_pn“, „mon_mitv_int“ und „Mon_uni_status_int“.

Aus der auf diese Art ermittelten Grundgesamtheit von 6.180.368 wird in der Folge eine 5%-Zufallsstichprobe gezogen. Das ergibt eine Stichprobengröße von n = 309.019.

³ Statistik Austria (2018), Abgestimmte Erwerbsstatistik und Erwerbsstatistik der Registerzählung 2011. Standard-Dokumentation Metainformationen (Definitionen, Erläuterungen, Methoden, Qualität). https://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&dDocName=040231

⁴ https://arbeitsmarktdatenbank.at/docs/Beschreibung_AM-Datenbank-Tabellen%20Uni_Status.pdf

⁵ Die Geburtsjahre 1955 bis 2005 umfassen allerdings einen Jahrgang mehr als die Altersjahre 15 bis 64. Insofern besteht in der AMDB eine geringfügige Untererfassung von rund 45000 Personen, was etwa einem halben Altersjahrgang (zu Jahresbeginn 2020 gab es 96157 14-Jährige und 85466 65-Jährige) bzw. 0,8% der 15- bis 64-Jährigen entspricht.

3 Erwerbsverlaufs-Datensatz (recov-ewv_v1)

3.1 Demografische Variablen (Block A)

Der Datentabelle „hv_pn“ wird das Geburtsjahr („gebj“) und das Geschlecht („gesl“) entnommen, wobei der zuletzt gültige Eintrag herangezogen wird. Diesem Eintrag wird auch der akademische Grad („fk_akgr“) entnommen; wenn kein akademischer Grad verzeichnet ist, wird ein allenfalls in früheren Einträgen vorhandener akademischer Grad herangezogen. Für jene, die bisher in Österreich noch nicht selbst versichert waren und daher noch kein Eintrag vorhanden ist, wird in der Datentabelle „mon_mitvers_int“ nach dem Geburtsjahr („geb_jahr_mitv“) und Geschlecht („gesl“) gesucht. Das Alter im Referenzjahr wird berechnet durch 2020 – gebj.

Der Datentabelle „mon_zusatz_bus“ wird der Wohnort des Hauptwohnsitzes entnommen, einmal in den 35 Einheiten der NUTS 3 Ebene (nuts, „Nomenclature des unités territoriales statistiques“) und einmal in der Kategorie der Regionalgeschäftsstellen des AMS („rgs“). Die Variable rgs enthält auch die Kategorie „AUSL“ für im Ausland wohnhafte Personen. Bei mitversicherten Personen entspricht der angegebene Wohnort dem Wohnort der Hauptversicherten. Bundesland und Typ der Region werden von der Variable „rgs“ abgeleitet. Typ der Region folgt einer Einstufung des AMS, wobei Großstädte, regionale Zentren, regionale Zentren mit hoher Dynamik, kleinere Regionen, kleinere Regionen mit hoher Dynamik und saisondominierte Regionen unterschieden werden.

Aus der Datentabelle „mon_mig_bus“ wird die Staatsangehörigkeit im Referenzjahr 2020 entnommen. In dieser Datentabelle ist bei Vorliegen einer ausländischen Staatsangehörigkeit diese eingetragen und im Falle einer erfolgten Einbürgerung die der österreichischen vorangegangenen Staatsangehörigkeit, einschließlich des Datums der Einbürgerung. Personen, die in dieser Datentabelle nicht erfasst sind, haben eine österreichische Staatsangehörigkeit ohne Wechsel der Staatsangehörigkeit. In diesem Fall wird als aktuelle Staatsangehörigkeit („nat_akt_kat3“) sowie als Staatsangehörigkeit davor („nat_herk_kat3“) jeweils Österreich angegeben. Bei einer erfolgten Einbürgerung ist „nat_akt_kat3“ ebenfalls Österreich und „nat_herk_kat3“ die unmittelbar vor dem Staatsbürgerschaftswechsel vorgelegene Staatsangehörigkeit. Bei Personen mit ausländischer Staatsangehörigkeit (wenn keine Einbürgerung vorliegt) ist diese Staatsangehörigkeit sowohl als „nat_akt_kat3“ als auch als „nat_herk_kat3“ eingetragen. Die Angabe erfolgt jeweils in einer groben geografischen Klassifikation (Österreich, EU-Länder mit Beitritt ab 2004 jedoch ohne Kroatien, EU-Länder vor 2004 inkl. EWR-Staaten, Großbritannien und Schweiz, Westbalkan-Länder inkl. Kroatien, Türkei, Staaten des Nahen und mittleren Ostens, einschließlich Maghreb (angelehnt an eine Version der MENA-Länder), sowie restliche Länder. Bei einer Einbürgerung ist die Variable „eingeb“ mit 1 codiert, ansonsten 0, und das Jahr der Einbürgerung in der Variablen „jahr_einb“ eingetragen. Die Variable „migstat“ gibt den Eintrag der Datentabelle „mon_mig_bus“ (Variable Migrationsstatus) mit den Ausprägungen (1 = AusländerIn, 2 = Einbürgerung, 3 = Mitversichertes Kind von MigrantInnen, sonst missing).

Die Variablen „geburt_1“ bis „geburt_n“ beinhalten das Jahr der ersten, zweiten, etc. in der Datentabelle „hv_epi_stamm_bas“ seit 1997 verzeichneten Lebendgeburt. Eingetragen wird das Ereignis Geburt, unabhängig davon, wie viele Kinder geboren wurden (im Fall von Zwilling- oder Mehrlingsgeburten erfolgt wie bei einer Einzelgeburt jeweils ein Eintrag). Es kann also nicht die Anzahl der Kinder, die eine Frau zwischen 1997 und Februar 2022 zur Welt gebracht hat, sondern die Zahl der Geburten in diesem Zeitraum, bei denen sie mindestens ein lebendiges Kind zu Welt gebracht hat, abgeleitet werden.

3.2 Arbeitsmarktstatus im Zeitverlauf (Block B)

Block B umfasst zunächst Variablen zu den in den Daten abgebildeten ersten Versicherungsperioden sowie die Gesamtdauer der verzeichneten Erwerbstätigkeits-, Arbeitslosigkeits- und Lehrlingsausbildungsperioden. Dann folgt der monatliche Arbeitsmarktstatus im Verlauf der Pandemie mit einem mehr als einjährigen Vorlauf von Jänner 2019 bis Dezember 2021. Diese Variablen basieren auf den Datentabellen „mon_uni_status_int“ und „hv_epi_stamm_bas“.

In der überschneidungsfrei aufbereiteten Datentabelle „mon_uni_status_int“ wird bei mehreren parallelen Episoden entsprechend der Hierarchisierung des AMS-Verbleibs- und Erwerbskarrieremonitorings vorgegangen. So hat z.B. laut der Dimensionstabelle „mon_uni_status_dim“ (siehe AMS - BMAFJ, 2020, S. 32–34 und Arbeitsblatt "B_uni_status" im Codebook zum Datensatz "recov-ewv_v1" bzw. "uni_status" im Codebook zum Datensatz "recov-eh_v1") eine geringfügige Beschäftigung einen niedrigeren Rang als Arbeitslosigkeit, AMS-Qualifikation oder AMS-Vormerkung und auch die gesicherten erwerbsfernen Positionen wie z.B. Karenz, Kinderbetreuung, Rente oder Wochengeld haben einen höheren Rang, weshalb bei Vorliegen einer solchen Qualifikation die geringfügige Beschäftigung in der Datentabelle „mon_uni_status_int“ nicht aufscheint. Die Angaben zur Gesamtdauer in geringfügigen Beschäftigungsverhältnissen (Kapitel 3.2.1) basieren deshalb auf der Datentabelle „hv_epi_stamm_bas“, die alle geringfügigen Beschäftigungsverhältnisse im DV-Universum darstellt, während die geringfügige Beschäftigung als monatlicher Arbeitsmarktstatus (Kapitel 3.2.2), als jährliche Statustage (Kapitel 3.3) sowie als Episoden (Kapitel 4) nur dann aufscheint, wenn nicht gleichzeitig eine höherrangige Versicherungsperiode besteht.

3.2.1 Erste Versicherungsperioden und Gesamtdauer ausgewählter Episoden

Die Daten zur ersten verzeichneten sozialversicherungsrechtlich erfassten Episode („sv_1“), der ersten Arbeitsmarktepisode („lf_1“), allfällig vorliegender Lehrlingsausbildungen („le_beginn“, „le_ende“, „le_dauer“ und „le_gef“), Arbeitslosigkeitsepisoden („al_beginn“, „al_ende“, „al_dauer“) und Erwerbstätigkeitsepisoden („et_beginn“, „et_ende“, „et_dauer“ und „et_gef“) kommen aus der Datentabelle „mon_uni_status_int“ und die Angaben zu geringfügigen Beschäftigungsverhältnissen („g1_beginn“, „g1_ende“, „g1_dauer“ und „g1_epi“) entstammen der Datentabelle „hv_epi_stamm_bas“. Diese Daten sind vollständig ab dem 1.1.1997 vorhanden, wobei grundsätzlich das tatsächliche Anfangsdatum von jenen Episoden verzeichnet ist, die in das Jahr 1997 hineinreichen. Der Beginn dieser Episoden ist linkszensiert, das früheste Anfangsdatum ist bei Arbeitslosigkeitsepisoden der 31.12.1970, bei Erwerbstätigkeitsepisoden der 31.12.1971 und bei geringfügiger Beschäftigung der 1.1.1994. Lehrlingsausbildungsperioden sind ab dem 1.1.1988 verfügbar wobei deren Anfangsdaten mit diesem Datum linkszensiert sind.

Die Enddaten sämtlicher Episoden sind in der derzeitigen Datenversion (April 2022) mit März 2022 rechtszensiert.

3.2.2 Monatlicher Arbeitsmarktstatus in der Pandemie 2019 bis 2021

Der monatliche Arbeitsmarktstatus ist als der längste Status im jeweiligen Monat entsprechend der überschneidungsfreien Datentabelle „mon_uni_status_int“ definiert. Haben mehrere Status gleich viele Statustage, so wird ein Status zufällig ausgewählt. Bei Episoden mit Status "AA" (Arbeiter und Angestellte) "LE" (Lehrlinge), "FD" (Freie Dienstnehmer) und "G1" (Geringfügig Beschäftigte) wird auf Basis der Datentabelle "hv_epi_stamm_bas" eine Zuordnung zu Arbeiter (AR, LEAR, FDAR, G1AR) bzw. zu Angestellten (AN, LEAN, FDAN, G1A) vorgenommen (siehe Codebook Datei „Codebook-recov-ewv_v1.xlsx“ für weitere Details).

Pro Monat enthält der Datensatz 7 Variablen:

- 1) Eine Laufnummer für den Dienstgeber des HV-Dienstgeberkontos („lfdg“)
- 2) Längster monatlicher Status (UNI-status erweitert, siehe Arbeitsblatt „B_uni-status“ im Codebook (Datei „Codebook-recov-ewv_v1.xlsx“), Spalte „status“)
- 3) Längster monatlicher Status zu Arbeitsmarkt-Typen zusammengefasst (siehe Arbeitsblatt „B_uni-status“ im Codebook (Datei „Codebook-recov-ewv_v1.xlsx“), Spalte „st_typ“)
- 4) Geförderter Status: Art des Vertragsverhältnisses, wenn 2) eine vom AMS oder SMS geförderte unselbstständige Beschäftigung (inkl. Kurzarbeit) oder Lehrlingsausbildung ist
- 5) NACE-Wirtschaftsbereich des Betriebes
- 6) Ort des Dienstgebers in der Kategorie der Regionalgeschäftsstellen des AMS
- 7) Ort des Dienstgebers nach NUTS 3 Ebene

3.3 Jährliche Statustage 2016 bis 2021 (Block C)

Block C stellt für die 13 möglichen Ausprägungen der zusammengefassten Arbeitsmarkt-Typen (siehe Kapitel 3.2.2, Ziffer 3) die jährliche Dauer in Tagen in diesen Status zur Verfügung. Unterschieden werden dabei folgende Typen (siehe Arbeitsblatt „B_uni-status“ im Codebook Datei „Codebook-recov-ewv_v1.xlsx“, Spalte „st_typ“):

- 1) Arbeiter
- 2) Angestellte (einschl. Vertragsbedienstete)
- 3) Beamte
- 4) Kurzarbeit
- 5) Sonstige Erwerbstätigkeit (vorwiegend Freie Dienstverträge)
- 6) Lehre in Arbeiterberufen
- 7) Lehre in Angestelltenberufen
- 8) Geförderte Lehrlingsausbildung (AMS/SMS)
- 9) Selbstständige
- 10) Geringfügig Beschäftigte
- 11) Arbeitslos
- 12) Erwerbsfern (einschließlich Meldelücken)
- 13) Missing (nach Todesfall, daher Einträge erst ab 2020)

3.4 Einkommen nach Einkommensarten 2016 bis 2021 (Block D)

Block D enthält Einkommensinformationen für die Jahre 2016 bis 2021, wobei zu beachten ist, dass es sich dabei um die Beitragsgrundlage handelt, d.h. die Einkommensinformation liegt für beitragspflichtige Erwerbstätigkeiten bis zur Höchstbeitragsgrundlage vor. Die AMDB-Datentabelle „hv_beitrg“ stellt die beitragspflichtigen Versicherungstage in der Höhe der Beitragsgrundlage pro Person und Dienstgeber zur Verfügung. Das Feld „BMG“ enthält das Jahresbruttoeinkommen beim jeweiligen Dienstgeber ohne Sonderzahlungen bis zur Höchstbeitragsgrundlage, das Feld „SZ“ enthält die Sonderzahlungen. Die Einkommensinformation ist also nach oben zensiert mit der Höchstbeitragsgrundlage zur Sozialversicherung (siehe Tabelle 2). Nicht bzw. nur teilweise enthalten sind Einkünfte aus geringfügigen Beschäftigungsverhältnissen.⁶

⁶ Enthalten sind sie etwa dann, wenn einzelne Erwerbstätige Einkünfte aus mehreren parallelen geringfügigen Beschäftigungsverhältnissen beziehen, die in Summe die Geringfügigkeitsgrenze übersteigen (Tabelle 2).

Tabelle 2: Höchstbeitragsgrundlagen und Geringfügigkeitsgrenze 2016 bis 2021

Jahr	Höchstbeitrags- grundlage (BMG)	Sonderzahlungen (SZ)	Bruttobezug gesamt (BMG + SZ)	Geringfügigkeits- grenze
2016	68040	9720	77760	4989
2017	69720	9960	79680	5108
2018	71820	10260	82080	5257
2019	73080	10440	83520	5362
2020	75180	10740	85920	5528
2021	77700	11100	88800	5710

Im Datensatz werden die Variablen „BMG“ und „SZ“ addiert, das „entspricht dem Bruttobezug gemäß ESTG 1988 §25 (ohne §26 und ohne Familienbeihilfe) der auf dem ‚Lohnzettel und Beitragsgrundlagennachweis‘ unter der Position 210 ausgewiesen wird“ (AMS - BMAFJ, 2020, S. 12). Beitragspflichtige Versicherungszeiten untergliedern sich in Zeiten einer Erwerbstätigkeit (Pflichtversicherung als Arbeiterin/Arbeiter und als Angestellte/Angestellter nach dem ASVG; als Selbstständige/Selbstständiger nach dem GSVG/FSVG; als Landwirtin/Landwirt nach dem BSVG), Zeiten einer Teilpflichtversicherung (Bezug von Arbeitslosengeld, Notstandshilfe und Umschulungsgeld; Bezug von Kranken-, Wochen-, Übergangs- und Rehabilitationsgeld; Zeiten der Kindererziehung, Familienhospizkarenz und Pflegekarenz; Zeiten des Präsenz- und Zivildienstes) sowie Zeiten einer freiwilligen Versicherung in der Pensionsversicherung.

Im Datensatz werden vier Einkommensarten unterschieden:

- 1) Einkommen aus **unselbstständiger Beschäftigung** (inkl. Sonderzahlungen): Summe aller Bruttobezüge eines Jahres lt. der AMDB-Datentabelle „hv_beitrg“ wobei nur jene Dienstgeberkonten berücksichtigt werden, bei denen in den entsprechenden Jahren unselbstständige Beschäftigungsepisoden der Personen (lt. der AMDB-Datentabelle hv_epi_stamm_bas) registriert sind. Siehe Codebook für weitere Informationen. Einkommen aus geringfügiger unselbstständiger Beschäftigung und Zivildienst sind nur teilweise erfasst.
- 2) Einkommen aus **selbstständiger Beschäftigung** (inkl. Sonderzahlungen) die der Versicherungspflicht unterliegen bzw. nicht davon ausgenommen sind: Summe aller Bruttobezüge eines Jahres lt. der AMDB-Datentabelle „hv_beitrg“ wobei nur jene Dienstgeberkonten berücksichtigt werden, denen in den entsprechenden Jahren die selbstständige Beschäftigung (lt. der AMDB-Datentabelle „hv_epi_stamm_bas“) zugeordnet ist. Siehe Codebook für weitere Informationen.
- 3) **AMS-Bezüge** (Arbeitslosengeld, Notstandshilfe, Überbrückungshilfe, Bezug eines Pensionsvorschusses, Bezug einer Sonderunterstützung); AMS-Bezüge sind seit 2019 in der AMDB-Datentabelle hv_beitrg abgebildet.
- 4) Transfers aus den **Sozialbudgets** (Sozialhilfe, Mindestsicherung)

Für jede Einkommensart ist die Zahl der Beitragstage zur Sozialversicherung aus diesen Bezügen pro Jahr dargestellt und daraus abgeleitet der Tagsatz in Euro pro Beitragstag berechnet.

Zusätzlich ist die Summe aus diesen vier Einkommensarten im Datensatz abgebildet, wobei zu beachten ist, dass AMS-Bezüge ab 2019, nicht aber für die Jahre 2016 bis 2018, verfügbar sind. Für diese Einkommensarten sind im Datensatz ebenfalls die jährlichen Beitragstage inkludiert und daraus abgeleitet auch der Tagsatz je Einkommensart.

4 Episodendatensatz (recov-eh_v1)

Der Episodendatensatz stellt pro Person sämtliche sozialversicherungsrechtlichen Beschäftigungs-, Arbeitslosigkeits- und Nichterwerbsepisoden aggregiert nach Arbeitsmarkt-Typen seit 2016 bereit. Die Anzahl der Einträge (Zeilen) pro Person entspricht der Anzahl der aufeinander folgenden Episodentypen. Eine Episode ist jeweils mit Beginn- und Enddatum versehen.

Die Episoden werden analog zu Kapitel 3.2.2 und 3.3 zunächst auf Basis der überschneidungsfreien Datentabelle „mon_uni_status_int“ ermittelt und bei Episoden mit Status "AA" (Arbeiter und Angestellte) sowie "LE" (Lehrlinge) wird auf Basis der Datentabelle "hv_epi_stamm_bas" eine Zuordnung zu Arbeiter (AR, LEAR) bzw. zu Angestellten (AN, LEAN) vorgenommen. Die Episoden werden zu Kategorien zusammengefasst (siehe Spalte „st_typ“ im Arbeitsblatt „uni-status“ im Codebook, sowie zur Aufteilung von Status „AA“ Arbeitsblatt „hv_code“). Die Variable „st_typ_fb“ gibt die Art des Vertragsverhältnisses, wenn „st_typ“ eine Kurzarbeitsepisode oder eine vom AMS oder SMS geförderte Lehrlingsausbildung ist.

Verschiedene aufeinanderfolgende Episoden in „mon_uni_status_int“, die dem gleichen Arbeitsmarkt-Typ angehören (z.B. zwei unmittelbar aufeinanderfolgenden Angestellten-Episoden bei verschiedenen Betrieben) werden zu einer Episode zusammengefasst, wobei sie dem Betrieb mit längerer Episodendauer zugeordnet wird (bei gleich langen Episodendauer erfolgt eine zufällige Zuordnung).

Der Datensatz hat 1.613.296 Episoden (Zeilen) und 6 Variablen (Spalten). Jeder Person ist die eindeutige Laufnummer („lfpe“) zugeordnet, die eine Verknüpfung mit dem Datensatz „recov-ewv_v1“ ermöglicht. Die Episoden sind nach Verfügbarkeit mit der zugeordneten eindeutigen Laufnummer des Betriebes („lfdg“) versehen. Dies ermöglicht eine Verknüpfung mit den Betriebsinformationen aus dem Datensatz „recov-ewv_v1“ (NACE-Wirtschaftsbereich und Betriebsort, siehe Kapitel 3.2.2) sowie mit den darüber hinaus gehenden Informationen des Betriebsdatensatzes (Kapitel 5).

5 Betriebsdatensatz (recov-dg_v1)

Der Betriebsdatensatz stellt für den Zeitraum 2016 bis 2021 auf jährlicher Basis verschiedene zusätzliche Informationen auf Betriebsebene bereit, die anhand der eindeutigen Betriebsnummer „lfdg“ (Laufnummer Dienstgeber) mit den anderen Datensätzen verknüpfbar ist.

5.1 Jährliche Indikatoren 2016 bis 2021

5.1.1 Lohnsumme und Durchschnittslohn pro Tag und Beschäftigte

Der Betriebsdatensatz stellt die Summe der über das Dienstgeberkonto des Betriebes an seine Beschäftigten in einem Jahr ausbezahlte Beitragsgrundlage (bezeichnet als Lohnsumme) und daraus abgeleitet die durchschnittliche Beitragsgrundlage pro Tag und Beschäftigte (Lohnsumme dividiert durch die Summe der Beitragstage) bereit. Es gilt als wesentliche Einschränkung zu beachten, dass die in den Daten verzeichneten Einkommen nach oben mit der Höchstbeitragsgrundlage zur Sozialversicherung zensiert sind (siehe Kapitel 3.4 und Tabelle 2). Außerdem sind Einkommen aus geringfügigen Beschäftigungsepisoden nicht oder nur zum Teil abgebildet. Diese sind in der abgeleiteten Variablen „Durchschnittslohn pro Tag und Beschäftigte“ nur insoweit enthalten, als Beitragsgrundlagen und Beitragstage verfügbar sind.

Datengrundlage und methodisches Vorgehen im Hinblick auf das Einkommen ist äquivalent zu Kapitel 3.4.

5.2 Versicherungsträger, Standort und NACE-Sektor

Auf jährlicher Basis ist der für den Betrieb zuständige Sozialversicherungsträger (Variable „vstr_...“), die AMS-Regionalgeschäftsstelle („dg_rgs_...“) sowie der NUTS3-Code („dg_nuts_...“) des angegebenen Betriebsstandortes und der NACE Wirtschaftsbereich („nace_...“) des Dienstgebers, jeweils analog zu Kapitel 3.2.2, enthalten.

5.3 Monatliche Betriebsindikatoren

Zum monatlichen Stichtag (jeweils Ende des Monats) stellt der Datensatz außerdem fünf weitere Indikatoren des Dienstgebers für den Zeitraum Januar 2016 bis Dezember 2021 bereit:

- 1) Frauenanteil
- 2) Durchschnittsalter (Median)
- 3) Ausländeranteil (Anteil mit nicht-österreichischer Staatsbürgerschaft)
- 4) Personalstand
- 5) Anzahl geringfügig Beschäftigter

Datenquelle ist die aufbereitete Datentabelle „mon_betrieb_bus“ (Betriebsmonitoring, AMS - BMAFJ, 2020, S. 42–46).

6 Materialien und Ressourcen

AMS - BMAFJ. (2020, Februar 5). Arbeitsmarktdatenbank (AMDB) - Tabellenbeschreibung UNI-Status. Verfügbar unter: https://arbeitsmarktdatenbank.at/docs/Beschreibung_AM-Datenbank-Tabellen%20Uni_Status.pdf

Codebooks:

Codebook-recov-ewv_v1.xlsx

Codebook-recov-eh_v1.xlsx

Codebook-recov-dg_v1.xlsx

Wiki IHS-DSC: <https://dsc-dok.ihs.ac.at/index.php/Arbeitsmarktdatenbankarbeit>

<https://arbeitsmarktdatenbank.at/>