

## Snabbguide till matriserna

- Matriser med *skill relatedness* data finns för fyra olika versioner av Svensk Näringsgrensindelning (SNI): SNI 2007, SNI 2002, SNI 92 och SNI 69.
- För varje version av SNI finns i sin tur **en matris för varje SNI-nivå**.
- Varje matris innehåller **en rad per möjlig kombination av SNI-koder**. I SNI 2007 på femsiffernivån finns exempelvis 821 olika bransch-koder, matrisen har därför  $821^2 = 674\,041$  rader.
- För varje kombination finns ett *relatedness index*, som indikerar hur relaterade de två branscherna är. Indexet sträcker sig mellan -1 och 1, där 1 är den högsta möjliga graden av skill relatedness.
- **Metoden** för att beräkna skill relatedness utifrån flöden av arbetare mellan branscher är hämtad från Neffke, Otto och Weyh (2017), och beskrivs i rapporten. Högt indexvärde betyder att det fanns ett relativt stort flöde av arbetare mellan branscherna.
- **Indexet är oriktat**, vilket innebär att för en given kombination av två branscher så är indexet detsamma oavsett vilken av branscherna som är bransch *i* respektive bransch *j*. Varje bransch-kombination förekommer alltså två gånger i matrisen.
- I många fall var det för få arbetare som bytte jobb mellan två branscher för att det skulle gå att beräkna ett index. Dessa tilldelas indexvärdet -1. Branscher definieras som perfekt relaterade till sig själva (index=1).
- Det är en ganska stor andel av bransch-kombinationerna som har fått indexvärdet -1, och i synnerhet för mindre branscher. Det understryker att matriserna i första hand är användbara för att identifiera minoriteten starka relationer, snarare än att särskilja nyansskillnader mellan det stora flertalet svagare relationer.
- **Vissa branscher observeras inte** överhuvudtaget i Sverige (exempelvis *odling av ris*), eller var så små att det inte gick att beräkna index till någon annan bransch. Dessa branscher finns med i matriserna, men har inga indexvärden, eftersom vi inte kan säga något om deras relatedness utifrån svenska data.
- Huvud-indexet (*rel\_index*) baseras på samtliga arbetare. Därutöver inkluderas **två alternativa index**, som baserar sig på arbetare med hög utbildningsnivå (*rel\_index\_hogutb*) respektive hög inkomst (*rel\_index\_hogink*).
- Matriserna erbjuds i två format: **Stata-format (.dta)** och **text-format (.csv)**. Kolumnerna i text-versionen avgränsas av tabbar. Csv är ett format som kan importeras till i stort sett alla program för databehandling, såsom Excel, R, Stata, SQL eller Python. För att svenska tecken ska läsas in korrekt när man importerar till exempelvis Excel kan man behöva ange teckenkodningen UTF-8.
- Vissa nivåer av SNI, exempelvis *avdelningsnivån* i SNI 2007, använder **bokstavskoder** i stället för sifferkoder. I dessa fall har varje bokstav ersatts av en siffra enligt alfabetisk ordning, det vill säga A=1, B=2, och så vidare.