Supplementary Material

Model-based estimation of small area dissimilarity indexes: An application to sex occupational segregation in Spain[1]

May 17, 2024

A Artificial data simulation experiments

Under the three-fold Fay-Herriot (FH3) model, two simulation experiments were carried out and both the explanatory and dependent variables were artificially generated. They investigate the behavior of the residual maximum likelihood (REML) estimators of the model parameters, the predictors of the Duncan segregation indexes (DSI) and the bootstrap estimators of the mean squared error (MSE). For $d = 1,\dots,D$, $r = 1,\dots,R$ ($R = 10$), $t = 1,\dots,T$ ($T = 5$), the set of explanatory variables and true parameters of the theoretical scenarios were simulated according to the following expressions:

$$
\begin{aligned}
&x_{drt} = \frac{1}{5}(b_{drt} - a_{drt})U_{drt} + a_{drt},\ U_{drt} = \frac{t}{T+1},\ a_{drt} = 1,\\
&b_{drt} = 1 + \frac{1}{DR}\big((d-1)RT + (r-1)T + t\big),\\
&y_{drt} = \beta_1 + \beta_2 x_{drt} + u_{1,d} + u_{2,drt} + u_{3,drt} + e_{drt},\ \beta_1 = \frac{1}{30},\ \beta_2 = \frac{1}{3},\\
&c_{drt} = 2.01 + 0.1 z_{drt},\ N_{drt} = c_{drt}(3\cdot 10^4\beta_1 + 3\cdot 10^4\beta_2 x_{drt}),
\end{aligned}
$$

where $u_{1,d} \sim N(0,\sigma_1^2)$ with $\sigma_1^2 = 0.001$, $u_{2,dr} \sim N(0,\sigma_2^2)$ with $\sigma_2^2 = 0.001$, $u_{3,drt} \sim N(0,\sigma_2^2)$ with $\sigma_3^2 = 0.001$, $z_{drt} \sim N(0,1)$ and $e_{drt} \sim N(0,\sigma_{drt}^2)$ with

$$
\sigma_{drt}^2 = \left(8\cdot\left(\frac{(d-1)RT + (r-1)T + t - 1}{DRT-1}\right) + 0.8\right)\cdot 10^{-6}.
$$

It was assumed that the dependent variable $y_{drt} = \hat{\overline{Y}}{}_{drt}^{dir}$ follows the FH3 model defined in (3.5). These are domain means of dichotomous variables, not totals, so they must take values between 0 and 1. Hence, simulations based on artificial data are restricted to a parametric scenario with limited variability. The latter has a significant influence on the absolute and relative error measures, with unusually low values compared to the simulation scenario based on real data.

A.1 Simulation A.1

Simulation A.1 has the following steps.

1. Repeat $I = 10^3$ times ($i = 1,\dots,I$):

[1]

1.1. For $d=1,\dots,D$, $r=1,\dots,R$, $t=1,\dots,T$, generate $(x_{drt}, N_{drt}, y_{drt}^{(i)})$ and calculate $\mu_{drt}^{(i)} = \beta_1 + \beta_2 x_{drt} + u_{1,d}^{(i)} + u_{2,dr}^{(i)} + u_{3,drt}^{(i)}$ and

$$S_{d.t}^{(i)} = \frac{1}{2}\sum_{r=1}^{R} S_{drt}^{(i)}, \quad S_{drt}^{(i)} = \left| \frac{N_{drt}\mu_{drt}^{(i)}}{\sum_{i=1}^{R} N_{dit}\mu_{dit}^{(i)}} - \frac{N_{drt}(1-\mu_{drt}^{(i)})}{\sum_{i=1}^{R} N_{dit}(1-\mu_{dit}^{(i)})} \right|.$$

1.2. Calculate the REML estimate $\hat{\tau}^{(i)} \in \{\hat{\beta}_1^{(i)}, \hat{\beta}_2^{(i)}, \hat{\sigma}_1^{2(i)}, \hat{\sigma}_2^{2(i)}, \hat{\sigma}_3^{2(i)}\}$.

1.3. For $d=1,\dots,D$, $r=1,\dots,R$, $t=1,\dots,T$, obtain the empirical best linear unbiased predictors (EBLUP) $\hat{\mu}_{drt}^{(i)}$ and the plug-in predictor of $S_{d.t}^{(i)}$, i.e.

$$\hat{S}_{d.t}^{in(i)} = \frac{1}{2}\sum_{r=1}^{R} \hat{S}_{drt}^{in(i)}, \quad \hat{S}_{drt}^{in(i)} = \left| \frac{N_{drt}\hat{\mu}_{drt}^{(i)}}{\sum_{i=1}^{R} N_{dit}\hat{\mu}_{dit}^{(i)}} - \frac{N_{drt}(1-\hat{\mu}_{drt}^{(i)})}{\sum_{i=1}^{R} N_{dit}(1-\hat{\mu}_{dit}^{(i)})} \right|.$$

1.4. Calculate the empirical marginal predictor (EMP) of $S_{d.t}^{(i)}$, denoted as $S_{d.t}^{emp(i)}$, by applying the antithetical Monte Carlo algorithm presented in Section 3.2.

2. For $\hat{\tau} \in \{\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2\}$ and $\hat{S}_{d.t} \in \{\hat{S}_{d.t}^{in}, \hat{S}_{d.t}^{emp}\}$, $d=1,\dots,D$, $t=1,\dots,T$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\tau}^{(i)} - \tau), \quad RMSE(\hat{\tau}) = \left(\frac{1}{I}\sum_{i=1}^{I}(\hat{\tau}^{(i)} - \tau)^2\right)^{1/2},$$

$$BIAS_{dt} = \frac{1}{I}\sum_{i=1}^{I}(\hat{S}_{d.t}^{(i)} - S_{d.t}^{(i)}), \quad RMSE_{dt} = \left(\frac{1}{I}\sum_{i=1}^{I}(\hat{S}_{d.t}^{(i)} - S_{d.t}^{(i)})^2\right)^{1/2},$$

$$ABIAS = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}|BIAS_{dt}|, \quad RMSE = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}RMSE_{dt}.$$

3. Calculate also the corresponding relative performance measures (in %), i.e.

$$RBIAS(\hat{\tau}) = 100\frac{BIAS(\hat{\tau})}{\tau}, \quad RRMSE(\hat{\tau}) = 100\frac{RMSE(\hat{\tau})}{\tau},$$

$$RBIAS_{dt} = 100\frac{BIAS_{dt}}{S_{d.t}}, \; RRMSE_{dt} = 100\frac{RMSE_{dt}}{S_{d.t}}, \; S_{d.t} = \frac{1}{I}\sum_{i=1}^{I} S_{d.t}^{(i)},$$

$$ARBIAS = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}|RBIAS_{dt}|, \quad RRMSE = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}RRMSE_{dt}.$$

Move D along $\{50, 100, 200, 300\}$. Tables A.1 and A.2 present the performance measures of the REML estimators of the model parameters, showing that the bias is always close to zero and the MSE decreases as D increases. Both bias and MSE are, in relative terms, more than acceptable. REML estimators seem to be asymptotically unbiased and consistent.

Table A.3 presents the performance measures of the DSI predictors. The bias remains stable as D increases for both predictors and the same applies to the MSE. Nevertheless, this behaviour is to be expected: as the number of small areas increases, so does the number of terms to predicted. Simulation A.1 shows that the plug-in predictor performs better than the EMP in terms of bias and MSE. These results agree with those obtained in the real data simulation experiment.

Measure	$D = 50$	$D = 100$	$D = 200$	$D = 300$
$BIAS(\hat{\beta}_1)$	0.0000010	-0.0000046	-0.0000572	0.0000570
$BIAS(\hat{\beta}_2)$	0.0000301	0.0000256	0.0000357	-0.0000114
$BIAS(\hat{\sigma}_1^2)$	-0.0000008	-0.0000004	0.0000001	0.0000005
$BIAS(\hat{\sigma}_2^2)$	0.0000002	-0.0000001	-0.0000007	-0.0000003
$BIAS(\hat{\sigma}_3^2)$	-0.0000001	-0.0000004	0.0000002	0.0000000
$RMSE(\hat{\beta}_1)$	0.0073816	0.0052606	0.0037258	0.0030759
$RMSE(\hat{\beta}_2)$	0.0045483	0.0032674	0.0023069	0.0019003
$RMSE(\hat{\sigma}_1^2)$	0.0002274	0.0001595	0.0001128	0.0000914
$RMSE(\hat{\sigma}_2^2)$	0.0000803	0.0000576	0.0000399	0.0000330
$RMSE(\hat{\sigma}_3^2)$	0.0000318	0.0000222	0.0000161	0.0000130

Table A.1: Absolute performance measures of the REML estimators.

Measure	$D = 50$	$D = 100$	$D = 200$	$D = 300$
$RBIAS(\hat{\beta}_1)$	0.0029	-0.0138	-0.1715	0.1710
$RBIAS(\hat{\beta}_2)$	0.0090	0.0077	0.0107	-0.0034
$RBIAS(\hat{\sigma}_1^2)$	-0.0795	-0.0428	0.0150	0.0464
$RBIAS(\hat{\sigma}_2^2)$	0.0226	-0.0139	-0.0724	-0.0337
$RBIAS(\hat{\sigma}_3^2)$	-0.0079	-0.0403	0.0198	-0.0049
$RRMSE(\hat{\beta}_1)$	22.1448	15.7819	11.1774	9.2277
$RRMSE(\hat{\beta}_2)$	1.3645	0.9802	0.6921	0.5701
$RRMSE(\hat{\sigma}_1^2)$	22.7422	15.9464	11.2848	9.1382
$RRMSE(\hat{\sigma}_2^2)$	8.0301	5.7640	3.9883	3.3005
$RRMSE(\hat{\sigma}_3^2)$	3.1766	2.2215	1.6119	1.2951

Table A.2: Relative performance measures of the REML estimators.

Predictor	Measure	$D = 50$	$D = 100$	$D = 200$	$D = 300$
Plug-in	$ABIAS$	0.0001001	0.0001007	0.0001008	0.0001017
EMP	$ABIAS$	0.0161594	0.0163307	0.0163246	0.0163352
Plug-in	$ARBIAS$	0.0634972	0.0640168	0.0638960	0.0642336
EMP	$ARBIAS$	12.0353598	12.1638334	12.1550223	12.1604527
Plug-in	$RMSE$	0.0026505	0.0026498	0.0026494	0.0026494
EMP	$RMSE$	0.0358118	0.0359969	0.0360153	0.0360437
Plug-in	$RRMSE$	1.5885767	1.5885171	1.5883617	1.5887864
EMP	$RRMSE$	24.9036768	25.0644015	25.0756856	25.0855677

Table A.3: Performance measures of the predictors.

A.2 Simulation A.2

Simulation A.2 studies the behavior of the MSE estimator of each DSI predictor, i.e, investigates the performance of $mse^*(\hat{S}^{in}_{d.t})$. To this end, such estimators are compared with the empirical MSE of $\hat{S}_{d.t}$, obtained from Simulation A.1. For ease of exposition, and to be closer to the real data-based scenario, we only take $D = 50$. It is important to note that, in the Spanish Labour Force Survey (SLFS) data, the number of provinces is $D = 52$.
The simulation procedure is as follows.

1. Set $D = 50$ and take $MSE_{dt} = RMSE_{dt}^2$ from the output of Simulation A.1.

2. Repeat $I = 500$ times ($i = 1, \ldots, I$):

   2.1. Generate $(x_{drt}, N_{drt}, y_{drt}^{(i)})$, $d = 1, \ldots, D$, $r = 1, \ldots, R$, $t = 1, \ldots, T$.

   2.2. Calculate the REML estimates $\hat{\beta}_1^{(i)}$, $\hat{\beta}_2^{(i)}$, $\hat{\sigma}_1^{2(i)}$, $\hat{\sigma}_2^{2(i)}$ and $\hat{\sigma}_3^{2(i)}$.

   2.3. Repeat B times ($b = 1, \ldots, B$):

   2.3.1. Generate $u_{1,d}^{*(ib)}$, $u_{2,dr}^{*(ib)}$, $u_{3,drt}^{*(ib)}$, $d = 1, \ldots, D$, $r = 1, \ldots, R$, $t = 1, \ldots, T$, by using $\hat{\sigma}_1^{2(i)}$, $\hat{\sigma}_2^{2(i)}$ and $\hat{\sigma}_3^{2(i)}$, instead of σ_1^2, σ_2^2 and σ_3^2. Generate $e_{drt}^{*(ib)} \sim N(0, \sigma_{drt}^2)$.

   2.3.2. Generate a bootstrap sample $\{y_{drt}^{*(ib)}; d = 1, \ldots, D, r = 1, \ldots, R, t = 1, \ldots, T\}$,

$$y_{drt}^{*(ib)} = \hat{\beta}_1^{(i)} + \hat{\beta}_2^{(i)} x_{drt} + u_{1,d}^{*(ib)} + u_{2,dr}^{*(ib)} + u_{3,drt}^{*(ib)} + e_{drt}^{*(ib)}.$$

   2.3.3. Calculate $\mu_{drt}^{*(ib)} = \hat{\beta}_1^{(i)} + \hat{\beta}_2^{(i)} x_{drt} + u_{1,d}^{*(ib)} + u_{2,dr}^{*(ib)} + u_{3,drt}^{*(ib)}$ and

$$S_{d.t}^{*(ib)} = \frac{1}{2} \sum_{r=1}^{R} S_{drt}^{*(ib)}, \quad S_{drt}^{*(ib)} = \left| \frac{N_{drt} \mu_{drt}^{*(ib)}}{\sum_{i=1}^{R} N_{dit} \mu_{dit}^{*(ib)}} - \frac{N_{drt}(1 - \mu_{drt}^{*(ib)})}{\sum_{i=1}^{R} N_{dit}(1 - \mu_{dit}^{*(ib)})} \right|.$$

   2.3.4. Calculate the REML bootstrap estimates $\hat{\beta}_1^{*(ib)}$, $\hat{\beta}_2^{*(ib)}$, $\hat{\sigma}_1^{2*(ib)}$, $\hat{\sigma}_2^{2*(ib)}$ and $\hat{\sigma}_3^{2*(ib)}$.

   2.3.5. For $d = 1, \ldots, D$, $r = 1, \ldots, R$, $t = 1, \ldots, T$, calculate the EBLUP

$$\hat{\mu}_{drt}^{*(ib)} = \hat{\beta}_1^{*(ib)} + \hat{\beta}_2^{*(ib)} x_{drt} + \hat{u}_{1,d}^{*(ib)} + \hat{u}_{2,dr}^{*(ib)} + \hat{u}_{3,drt}^{*(ib)}$$

   and the plug-in predictor of $S_{d.t}^{*(ib)}$, i.e.

$$\hat{S}_{d.t}^{*in(ib)} = \frac{1}{2} \sum_{r=1}^{R} \hat{S}_{drt}^{*in(ib)}, \ \hat{S}_{drt}^{*in(ib)} = \left| \frac{N_{drt} \hat{\mu}_{drt}^{*(ib)}}{\sum_{i=1}^{R} N_{dit} \hat{\mu}_{dit1}^{*(ib)}} - \frac{N_{drt}(1 - \hat{\mu}_{drt}^{*(ib)})}{\sum_{i=1}^{R} N_{dit}(1 - \hat{\mu}_{dit}^{*(ib)})} \right|.$$

   2.4 For $d = 1, \ldots, D$, calculate

$$mse_{dt}^{*(i)} = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{S}_{d.t}^{*in(ib)} - S_{d.t}^{*(ib)} \right)^2.$$

3. For $d = 1, \ldots, D$, calculate the absolute performance measures:

$$B_{dt} = \frac{1}{I} \sum_{i=1}^{I} \left(mse_{dt}^{*(i)} - MSE_{dt} \right), \ RE_{dt} = \left(\frac{1}{I} \sum_{i=1}^{I} \left(mse_{dt}^{*(i)} - MSE_{dt} \right)^2 \right)^{1/2}.$$

4. For $d = 1, \ldots, D$, $t = 1, \ldots, T$, calculate the relative performance measures (in %):

$$RB_{dt} = \frac{100 B_{dt}}{MSE_{dt}}, \ RRE_{dt} = \frac{100 RE_d}{MSE_{dt}},$$

$$ARB = \frac{1}{DT} \sum_{d=1}^{D} \sum_{t=1}^{T} |RB_{dt}|, \ RRE = \frac{1}{DT} \sum_{d=1}^{D} \sum_{t=1}^{T} RRE_{dt}.$$

Simulation A.2 aims to provide some advice on which B value to choose. Absolute measures are not interpretable because they depend on the magnitude of the target variable, i.e. the artificial proportion of employed men by estimation domain. This suggests focusing on the relative biases, RB_{dt}, and the relative root-MSEs, RRE_{dt}, $d = 1, \ldots, D$, $t = 1, \ldots, T$. For $B = 50, 100, 200, 300, 400$, Figure A.1 represents five boxplots of RB_{dt} and RRE_{dt}.
Firstly, the left boxplots in Figure A.1 show that the relative biases do not decrease as the size of B increases. Moreover, they behave symmetrically and close to zero, with a slightly negative but practically zero mean. Secondly, the right boxplots show that the relative root-MSEs decrease as B increases, achieving good results for values greater than or equal to 200 resamples. Thus, the conclusions obtained are in line with those presented in the scenario based on real data. However, the order of the results differ significantly because of the generation of the target population which, being a proportion, is forced to have values between 0 and 1.

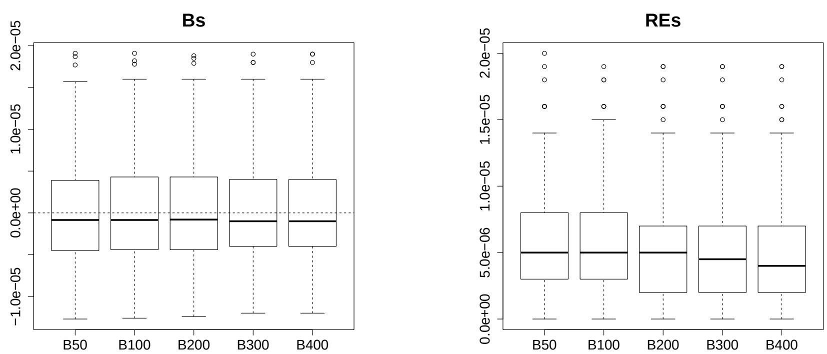

Figure A.1: Boxplots of RB_{dt}'s (left) and RRE_{dt}'s (right) for $B = 50, 100, 200, 300, 400$.

B Tables

Table B.1 presents the DSI predictions by Spanish province for the last quarter of 2020 and the four quarters of 2021. Data are from the SLFS. Analysing the temporal evolution, no major changes are observed. In terms of spatial dependencies, and as can be infered from Section 5, neighbouring provinces and those with similar demographic and socioeconomic conditions perform closer to each other. On average, the DSI predictions decrease from 0.2261 in 2020.4 to 0.2182 in 2021.4, with no significant changes at 5% and standard deviations close to 0.05.

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Álava	0.2907	0.2724	0.2263	0.2857	0.2702
Albacete	0.3071	0.2854	0.2830	0.2778	0.2910
Alicante	0.2339	0.2200	0.2284	0.2219	0.2105
Almería	0.2069	0.2669	0.2057	0.2289	0.2296
Ávila	0.1693	0.2048	0.1766	0.1609	0.1661
Badajoz	0.2459	0.2394	0.2616	0.2422	0.2408
Islas Baleares	0.2405	0.1703	0.2092	0.2355	0.2157
Barcelona	0.1833	0.1767	0.1777	0.1886	0.1865
Burgos	0.2193	0.2250	0.2064	0.2147	0.2181
Cáceres	0.2948	0.2774	0.2482	0.2910	0.2703

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Cádiz	0.2684	0.2532	0.2031	0.2735	0.2736
Castellón	0.2170	0.2028	0.1853	0.1919	0.2016
Ciudad Rea	0.2611	0.2407	0.2637	0.2289	0.2554
Córdoba	0.2311	0.2484	0.2421	0.2102	0.2475
A Coruña	0.1817	0.1682	0.1648	0.1529	0.1637
Cuenca	0.2125	0.2364	0.2164	0.2096	0.2178
Girona	0.1870	0.1846	0.1916	0.1834	0.1921
Granada	0.2247	0.2325	0.2365	0.2090	0.2166
Guadalajara	0.1759	0.1613	0.1525	0.1526	0.1439
Guipúzcoa	0.2202	0.2190	0.2010	0.1826	0.2015
Huelva	0.1928	0.1525	0.2133	0.2128	0.2201
Huesca	0.3195	0.2752	0.3326	0.2557	0.2762
Jaén	0.2825	0.3034	0.2310	0.2483	0.2657
León	0.1940	0.1824	0.2105	0.2310	0.2277
Lleida	0.2945	0.2938	0.2484	0.3085	0.2855
La Rioja	0.2001	0.1615	0.1676	0.2072	0.1759
Lugo	0.2624	0.2869	0.2636	0.2651	0.2808
Madrid	0.1942	0.2022	0.1886	0.2089	0.1914
Málaga	0.1521	0.1687	0.1678	0.1337	0.1496
Murcia	0.2480	0.2356	0.2329	0.2084	0.2276
Navarra	0.2102	0.2077	0.2007	0.2066	0.2186
Ourense	0.1670	0.1764	0.1731	0.1906	0.1651
Asturias	0.2136	0.2163	0.2267	0.2204	0.2130
Palencia	0.1953	0.2052	0.2151	0.1813	0.2034
Las Palmas	0.1595	0.1641	0.1279	0.1644	0.1665
Pontevedra	0.1733	0.1546	0.1518	0.1607	0.1640
Salamanca	0.2257	0.2184	0.1774	0.2288	0.2034
Tenerife	0.1849	0.1873	0.1917	0.2047	0.1860
Cantabria	0.1577	0.1587	0.1838	0.1876	0.1682
Segovia	0.1861	0.1543	0.1600	0.1928	0.1954
Sevilla	0.1913	0.1780	0.1709	0.1660	0.1833
Soria	0.2993	0.2681	0.2651	0.2820	0.2169
Tarragona	0.3054	0.2790	0.2561	0.2755	0.2554
Teruel	0.3503	0.3240	0.3230	0.3367	0.3407
Toledo	0.2780	0.2743	0.2787	0.2821	0.2759
Valencia	0.2032	0.2080	0.2094	0.1815	0.2165
Valladolid	0.1973	0.1941	0.1930	0.1716	0.1906
Vizcaya	0.2653	0.2391	0.2546	0.2080	0.2166
Zamora	0.2067	0.1929	0.1756	0.1543	0.1972
Zaragoza	0.1984	0.1992	0.2149	0.2037	0.1965
Ceuta	0.1668	0.1310	0.1547	0.1493	0.1556
Melilla	0.3107	0.2574	0.2818	0.2168	0.3052
Average	0.2261	0.2180	0.2139	0.2151	0.2182
SD	0.0494	0.0469	0.0441	0.0453	0.0440

Table B.1: DSI estimates. Data from the SLFS of 2020.4-2021.4.

Table B.2 presents the parametric bootstrap estimates of the RRMSE of the DSI predictors from Table B.1. Similar to the province-time period index predictions, once a province has been fixed, the error tends to remain stable over time. On average, the provincial RRMSE is close to 0.17 and its standard deviation exceeds 0.04 but does not reach 0.06.

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Álava	0.1349	0.1403	0.1703	0.1444	0.1506
Albacete	0.1297	0.1337	0.1373	0.1513	0.1459
Alicante	0.1503	0.1481	0.1451	0.1654	0.1676
Almería	0.1976	0.2019	0.2333	0.2174	0.2270
Ávila	0.2485	0.2018	0.2309	0.2749	0.2488
Badajoz	0.1532	0.1430	0.1430	0.1604	0.1551
Islas Baleares	0.1276	0.1575	0.1337	0.1343	0.1371
Barcelona	0.0942	0.1010	0.1037	0.1047	0.1051
Burgos	0.1762	0.1740	0.1767	0.1923	0.1800
Cáceres	0.1402	0.1305	0.1600	0.1441	0.1540
Cádiz	0.1327	0.1265	0.1669	0.1366	0.1332
Castellón	0.1461	0.1564	0.1783	0.1844	0.1722
Ciudad Rea	0.1363	0.1419	0.1356	0.1633	0.1442
Córdoba	0.1585	0.1544	0.1542	0.1818	0.1665
A Coruña	0.1191	0.1247	0.1293	0.1472	0.1395
Cuenca	0.2135	0.1806	0.2083	0.2363	0.2152
Girona	0.1794	0.1721	0.1816	0.2061	0.1886
Granada	0.1462	0.1394	0.1391	0.1759	0.1630
Guadalajara	0.2107	0.2337	0.2357	0.2725	0.2650
Guipúzcoa	0.1518	0.1525	0.1705	0.2013	0.2002
Huelva	0.2225	0.3031	0.2513	0.2116	0.2099
Huesca	0.1321	0.1423	0.1309	0.1745	0.1500
Jaén	0.1495	0.1379	0.1687	0.1662	0.1667
León	0.1913	0.1855	0.1729	0.1770	0.1732
Lleida	0.1430	0.1346	0.1599	0.1417	0.1474
La Rioja	0.1652	0.1928	0.1877	0.1593	0.2013
Lugo	0.1359	0.1222	0.1356	0.1422	0.1382
Madrid	0.0805	0.0797	0.0807	0.0774	0.0844
Málaga	0.2363	0.1981	0.2104	0.2783	0.2490
Murcia	0.1094	0.1113	0.1184	0.1448	0.1288
Navarra	0.1405	0.1426	0.1485	0.1543	0.1424
Ourense	0.2084	0.2054	0.2081	0.2028	0.2364
Asturias	0.1300	0.1294	0.1184	0.1328	0.1357
Palencia	0.2089	0.1919	0.1929	0.2453	0.2043
Las Palmas	0.2131	0.2099	0.2558	0.2172	0.1955
Pontevedra	0.1253	0.1428	0.1487	0.1469	0.1512
Salamanca	0.1847	0.1827	0.2222	0.1896	0.2011
Tenerife	0.1908	0.1782	0.1618	0.1734	0.1831
Cantabria	0.1972	0.1860	0.1604	0.1717	0.1821
Segovia	0.2154	0.2678	0.2441	0.2213	0.2125
Sevilla	0.1224	0.1326	0.1419	0.1569	0.1468
Soria	0.1489	0.1511	0.1515	0.1548	0.1914

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Tarragona	0.1181	0.1195	0.1341	0.1435	0.1431
Teruel	0.1243	0.1248	0.1275	0.1311	0.1305
Toledo	0.1274	0.1246	0.1261	0.1333	0.1383
Valencia	0.1114	0.1041	0.1114	0.1346	0.1144
Valladolid	0.1807	0.1793	0.1706	0.2077	0.1897
Vizcaya	0.1269	0.1330	0.1334	0.1681	0.1593
Zamora	0.2021	0.2262	0.2319	0.3008	0.2369
Zaragoza	0.1422	0.1447	0.1300	0.1496	0.1529
Ceuta	0.3125	0.3777	0.3496	0.3180	0.3097
Melilla	0.1581	0.1860	0.1813	0.2066	0.1693
Average	0.1616	0.1647	0.1692	0.1794	0.1737
SD	0.0439	0.0515	0.0481	0.0489	0.0430

Table B.2: RRMSE estimates of the DSI predictors. Data from the SLFS of 2020.4-2021.4.

According to the statistical results presented throughout the main document, the OC2 category has a substantially high contribution to the DSI. Moreover, it corresponds to those jobs occupied by scientific and intellectual technicians and professionals, which implied that its interpretation is particularly interesting. Therefore, two additional tables relating to the specific study of this category are included. First, Table B.3 presents the DSI-OC2 predictions by Spanish province for the last quarter of 2020 and the four quarters of 2021. On average, the DSI-OC2 predictions decrease from 0.1357 in 2020.4 to 0.1304 in 2021.4, with no significant changes at 5% and standard deviations close to 0.03. Again, neighbouring provinces and those with similar demographic and socioeconomic conditions behave similarly. Time dependence is evident.

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Álava	0.1330	0.1256	0.1059	0.1202	0.1348
Albacete	0.1619	0.1706	0.1843	0.1697	0.1966
Alicante	0.1254	0.1204	0.1232	0.1112	0.1091
Almería	0.1170	0.1672	0.1185	0.1339	0.1232
Ávila	0.1107	0.1190	0.1091	0.1041	0.1016
Badajoz	0.1471	0.1441	0.1487	0.1484	0.1348
Islas Baleares	0.1629	0.1052	0.1332	0.1342	0.1594
Barcelona	0.1189	0.1124	0.1128	0.1318	0.1222
Burgos	0.1210	0.1265	0.1110	0.1253	0.1166
Cáceres	0.2081	0.1988	0.1971	0.1862	0.2078
Cádiz	0.0939	0.0764	0.0846	0.1254	0.0769
Castellón	0.1100	0.1131	0.1005	0.0843	0.1109
Ciudad Rea	0.2051	0.1785	0.1855	0.1598	0.1634
Córdoba	0.1714	0.1656	0.1506	0.1501	0.1380
A Coruña	0.1241	0.1218	0.1116	0.0952	0.1038
Cuenca	0.1425	0.1424	0.1350	0.1126	0.1308
Girona	0.1286	0.1534	0.1362	0.1128	0.1237
Granada	0.1440	0.1240	0.1326	0.1429	0.1465
Guadalajara	0.0850	0.0770	0.0792	0.0912	0.0790
Guipúzcoa	0.0962	0.1174	0.1131	0.0834	0.1100
Huelva	0.1175	0.1034	0.1588	0.1001	0.0960

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Huesca	0.1852	0.1780	0.1977	0.1805	0.1619
Jaén	0.1788	0.1761	0.1695	0.1791	0.1825
León	0.1435	0.1341	0.1457	0.1566	0.1432
Lleida	0.1754	0.1833	0.1347	0.1894	0.1655
La Rioja	0.1302	0.1178	0.1076	0.1187	0.1136
Lugo	0.1263	0.1218	0.1377	0.1239	0.1361
Madrid	0.1216	0.1216	0.1178	0.1118	0.1352
Málaga	0.1233	0.1158	0.1239	0.0853	0.0989
Murcia	0.1748	0.1725	0.1542	0.1488	0.1424
Navarra	0.0839	0.0636	0.0730	0.0763	0.0868
Ourense	0.1506	0.1588	0.1379	0.1343	0.1477
Asturias	0.1236	0.1327	0.1256	0.1108	0.1241
Palencia	0.1251	0.1319	0.1489	0.1377	0.1464
Las Palmas	0.1146	0.1391	0.1230	0.1050	0.1136
Pontevedra	0.1122	0.0988	0.1081	0.0932	0.1149
Salamanca	0.1723	0.1634	0.1350	0.1483	0.1386
Tenerife	0.1465	0.1328	0.1244	0.1414	0.1541
Cantabria	0.1355	0.1036	0.1018	0.1078	0.1197
Segovia	0.1315	0.1258	0.1144	0.1279	0.1446
Sevilla	0.1289	0.1351	0.1220	0.1128	0.1194
Soria	0.1803	0.2040	0.1907	0.1387	0.1493
Tarragona	0.1322	0.1390	0.1371	0.1284	0.1165
Teruel	0.1328	0.1232	0.1163	0.1173	0.1424
Toledo	0.1222	0.1340	0.1248	0.1034	0.1160
Valencia	0.1139	0.1445	0.1183	0.1035	0.1360
Valladolid	0.0971	0.1255	0.0965	0.0761	0.1024
Vizcaya	0.1105	0.1338	0.1114	0.0805	0.0729
Zamora	0.0975	0.1150	0.1067	0.0768	0.1358
Zaragoza	0.0861	0.1154	0.1185	0.1145	0.1063
Ceuta	0.1429	0.1679	0.1183	0.1088	0.1140
Melilla	0.2350	0.2032	0.2091	0.1280	0.2161
Average	0.1357	0.1361	0.1304	0.1229	0.1304
SD	0.0325	0.0311	0.0300	0.0289	0.0302

Table B.3: DSI-OC2 estimates. Data from the SLFS of 2020.4-2021.4.

Table B.4 presents the parametric bootstrap estimates of the RRMSE of the DSI-OC2 predictors. On average, they range from 0.25 to 0.30, with values twice as high as those obtained for the DSI. The standard deviations are also larger, increasing the uncertainty of the RRMSE estimates. As already reasoned in Section 5, these values are justified by the fact that predicting an average value (the DSI) presents less uncertainty than predicting its summands (the DSI-OC2). Nevertheless, it can be concluded that the accuracy is still adequate for the individual prediction of a specific main occupation as far as a SAE problem is concerned.

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Álava	0.3103	0.3518	0.4175	0.3726	0.2765
Albacete	0.2391	0.2328	0.2235	0.2405	0.1736

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Alicante	0.2662	0.2662	0.2828	0.3048	0.2617
Almería	0.2286	0.2039	0.2575	0.2648	0.2230
Ávila	0.3072	0.3100	0.2988	0.3193	0.2883
Badajoz	0.2137	0.2034	0.2067	0.2283	0.1986
Islas Baleares	0.1776	0.2171	0.2039	0.1898	0.1657
Barcelona	0.1341	0.1536	0.1441	0.1416	0.1316
Burgos	0.2645	0.2771	0.2870	0.2719	0.2481
Cáceres	0.1841	0.1840	0.1910	0.1994	0.1599
Cádiz	0.3316	0.3825	0.3839	0.2765	0.3449
Castellón	0.2515	0.2569	0.2867	0.3363	0.2274
Ciudad Rea	0.1881	0.1963	0.2095	0.2418	0.1905
Córdoba	0.2228	0.2240	0.2461	0.2413	0.2197
A Coruña	0.1637	0.1719	0.1835	0.2281	0.1687
Cuenca	0.2643	0.2614	0.2917	0.3428	0.2457
Girona	0.2504	0.2312	0.2714	0.3324	0.2572
Granada	0.2213	0.2620	0.2301	0.2770	0.2083
Guadalajara	0.3934	0.3986	0.3568	0.3748	0.3450
Guipúzcoa	0.3879	0.3555	0.3725	0.4657	0.3284
Huelva	0.2837	0.3200	0.2491	0.3281	0.3062
Huesca	0.2056	0.2205	0.2083	0.2214	0.2090
Jaén	0.2013	0.2091	0.2199	0.2187	0.1744
León	0.2680	0.2642	0.2694	0.2924	0.2294
Lleida	0.2355	0.2139	0.2717	0.2182	0.2074
La Rioja	0.2334	0.2531	0.2916	0.2628	0.2346
Lugo	0.2432	0.2709	0.2519	0.3000	0.2237
Madrid	0.1224	0.1296	0.1300	0.1367	0.1053
Málaga	0.3023	0.3029	0.2961	0.3844	0.2966
Murcia	0.1458	0.1444	0.1704	0.1867	0.1551
Navarra	0.3850	0.4810	0.4394	0.4671	0.3093
Ourense	0.2287	0.2239	0.2549	0.2490	0.2087
Asturias	0.2329	0.2142	0.2225	0.2763	0.2048
Palencia	0.3019	0.2752	0.2747	0.2669	0.2229
Las Palmas	0.2495	0.2421	0.2419	0.2908	0.2234
Pontevedra	0.1688	0.1962	0.2026	0.2322	0.1639
Salamanca	0.2482	0.2596	0.3211	0.2873	0.2402
Tenerife	0.2322	0.2229	0.2207	0.2282	0.1769
Cantabria	0.2135	0.2689	0.2901	0.2783	0.2132
Segovia	0.2586	0.2671	0.2591	0.2532	0.1992
Sevilla	0.1630	0.1723	0.1897	0.2332	0.1775
Soria	0.1911	0.1637	0.1697	0.2521	0.1751
Tarragona	0.2576	0.2540	0.2611	0.3020	0.2666
Teruel	0.2819	0.2918	0.3083	0.3196	0.2395
Toledo	0.2516	0.2438	0.2470	0.2698	0.2274
Valencia	0.1694	0.1540	0.1807	0.2208	0.1442
Valladolid	0.3937	0.3365	0.3877	0.5295	0.3216
Vizcaya	0.3354	0.2850	0.3388	0.4668	0.4113
Zamora	0.3580	0.3219	0.3515	0.4924	0.2592

Province	2020.4	2021.1	2021.2	2021.3	2021.4
Zaragoza	0.3000	0.2430	0.2495	0.2735	0.2313
Ceuta	0.3584	0.3196	0.4033	0.3772	0.3242
Melilla	0.2939	0.3219	0.3359	0.3996	0.2523
Average	0.2522	0.2544	0.2664	0.2916	0.2307
SD	0.0674	0.0687	0.0706	0.0854	0.0611

Table B.4: RRMSE estimates of the DSI-OC2 predictors. Data from the SLFS of 2020.4-2021.4 in OC2 category.