

Regional Income Differences and Market Accessibility in the Enlarged European Union

A Simultaneous Spatial CGE and Estimation Approach

Nils Happich

Department of Economics
Faculty of Economics, Business and Social Sciences
Christian-Albrechts-University at Kiel

July 2009

Regional Income Differences and Market Accessibility in the Enlarged European Union

A Simultaneous Spatial CGE and Estimation Approach

Inaugural-Dissertation zur Erlangung des akademischen Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Diplom-Volkswirt Nils Happich
aus Wuppertal

Kiel, Juli 2009

Christian-Albrechts-Universität zu Kiel
Institut für Regionalforschung
Wilhelm-Seelig-Platz 1
24098 Kiel

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät der
Christian-Albrechts-Universität zu Kiel

Dekan: Prof. Dr. Thomas Lux

Erstberichterstatter:
Prof. Dr. Johannes Bröcker

Zweitberichterstatter:
Prof. Dr. Helmut Herwartz

Abgabedatum: 8. Juli 2009
Prüfungsdatum: 22. Februar 2010

TO MY FAMILY

Contents

List of Tables	V
List of Figures	VII
List of Country Abbreviations	IX
List of Regions	XI
1 Introduction	1
2 A Critical Literature Review	5
2.1 Theoretical Foundations	5
2.1.1 Monopolistic Competition Trade Models	6
2.1.2 New Economic Geography	10
2.2 Survey of Empircis	13
2.2.1 The Hanson Approach	14
2.2.2 The Redding-Venables Approach	19
2.2.3 Critique	24
2.3 Chapter Summary	26
3 The Analytical Framework	28
3.1 The <i>CGEurope Model</i>	29
3.1.1 Assumptions	29
3.1.2 Supply Side	29
3.1.3 Demand Side	31
3.1.4 Transportation and Trade	32
3.1.5 Equilibrium	34
3.2 The Empirical Specification	35
3.2.1 Calibration of the Model	35

3.2.2	The Factor Price Equation	37
3.3	Chapter Summary	39
4	Econometric Theory and Application	40
4.1	Cross-Section Specifications	40
4.1.1	Nonlinear Least Squares Approach	40
4.1.2	Instrumental Variables Estimation	43
4.1.3	Spatial Econometrics Problems	45
4.2	Panel Specifications	50
4.2.1	Nonlinear Least Squares Approach	50
4.2.2	Instrumental Variables Estimation	52
4.2.3	Spatial Econometrics Problems	53
4.3	Chapter Summary	53
5	Descriptive Data Analysis and Preprocessing	54
5.1	Data Requirements and Availability	54
5.2	Regional System	55
5.3	Descriptive Data Analysis	56
5.3.1	National Data	59
5.3.2	Regional Data	59
5.4	Data Preprocessing	66
5.4.1	National Data	66
5.4.2	Regional Data	69
5.5	Chapter Summary	72
6	Estimating the Wage Equation for the Enlarged European Union	73
6.1	Preliminaries	73
6.2	Cross-Section Results	75
6.2.1	NLS Specifications	77
6.2.2	Spatial Econometrics Models	81
6.2.3	Sensitivity Analysis	88
6.3	Panel Results	92
6.4	Chapter Summary	98
7	Conclusions	100
	Bibliography	102

Curriculum Vitae 113

Certificate of Authorship/Originality (Eidesstattliche Erklärung) 115

List of Tables

2.1	Key Parameter Estimates in Literature	24
5.1	Regional Entities	57
5.2	Comparability of Regional Entities	58
5.3	National Parameter Values	60
5.4	Data Sources	62
6.1	Cross-Section Results: NLS Specification 1	76
6.2	Cross-Section Results: NLS Specification 2	80
6.3	Cross-Section Results: Spatial Models	84
6.4	Cross-Section Results: Variation over Time	90
6.5	Cross-Section Results: Functional Sensitivity (Specification 1)	91
6.6	Panel Results: Random Effects Models	94

List of Figures

5.1	Regional GDP per Capita Levels (PPP) 1995	65
5.2	Regional GDP per Capita Growth from 1995 to 2002	67
5.3	Regional GDP per head (PPP) in Germany (1995)	68
5.4	Corrected regional GDP per Capita Levels (PPP) 1995	71
6.1	Distance & Transport Cost Functions (Cross-Section Specification 1)	78
6.2	Distance & Transport Cost Functions (Cross-Section Specification 2)	82
6.3	Distance & Transport Cost Functions (Spatial Cross-Section Models)	85
6.4	Accessibility Measures 2002 (Cross-Section Specification 1)	86
6.5	Accessibility Measures 2002 (Cross-Section Specification 2)	87
6.6	Functional Sensitivity of Distance & Transport Cost Functions (Spec- ification 1)	92
6.7	Functional Sensitivity of Distance & Transport Cost Functions (Spa- tial Model Specification 1)	92
6.8	Distance & Transport Cost Functions (Panel Models)	95
6.9	Accessibility Measures 2002 (Panel Specification 1)	96
6.10	Accessibility Measures 2002 (Panel Specification 2)	97

List of Country Abbreviations

AL	Albania
AT	Austria
BA	Bosnia
BE	Belgium
BG	Bulgaria
BY	Belarus
CH	Switzerland
CY	Cyprus
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GR	Greece
HR	Croatia
HU	Hungary
IE	Ireland
IS	Iceland
IT	Italy
LI	Liechtenstein
LT	Lithuania
LU	Luxembourg
LV	Latvia
MD	Moldova
MK	Former Yugoslavian Republic of Macedonia
ML	Malta

NL	Netherlands
NO	Norway
PL	Poland
PT	Portugal
RO	Romania
ROW	Rest of the world
RU	Russian Federation
SE	Sweden
SI	Slovenia
SK	Slovakia
TR	Turkey
UA	Ukraine
UK	United Kingdom
YU	Yugoslavian Republic of Serbia and Montenegro

List of Regions

1 Mittelburgenland, 2 Nordburgenland, 3 Südburgenland, 4 Mostviertel-Eisenwurzen, 5 Niederösterreich-Süd, 6 St Pölten, 7 Waldviertel, 8 Weinviertel, 9 Wiener Umland/Nordteil, 10 Wiener Umland/Südteil, 11 Wien, 12 Klagenfurt-Villach, 13 Oberkärnten, 14 Unterkärnten, 15 Graz, 16 Liezen, 17 Östliche Obersteiermark, 18 Oststeiermark, 19 West- und Südsteiermark, 20 Westliche Obersteiermark, 21 Innviertel, 22 Linz-Wels, 23 Mühlviertel, 24 Steyr-Kirchdorf, 25 Traunviertel, 26 Lungau, 27 Pinzgau-Pongau, 28 Salzburg und Umgebung, 29 Ausserfern, 30 Innsbruck, 31 Osttirol, 32 Tiroler Oberland, 33 Tiroler Unterland, 34 Bludenz-Bregenzer Wald, 35 Rheintal-Bodenseegebiet, 36 Bruxelles Capitale, 37 Antwerpen, 38 Mechelen, 39 Turnhout, 40 Hasselt, 41 Maaseik, 42 Tongeren, 43 Aalst, 44, Dendermonde, 45 Eeklo, 46 Gent-Arrondissement, 47 Oude-naarde, 48 Sint-Niklaas, 49 Halle-Vilvoorde, 50 Leuven, 51 Brugge, 52 Diksmuide, 53 Ieper, 54 Kortrijk, 55 Oostende, 56 Roeselare, 57 Tielt, 58 Veurne, 59 Brabant Wallon, 60 Ath, 61 Charleroi, 62 Mons, 63 Mouscron, 64 Soignies, 65 Thuin, 66 Tournai, 67 Huy, 68 Liège Arrondissement, 69 Verviers, 70 Waremme, 71 Arlon, 72 Bastogne, 73 Marche-En-Famenne, 74 Neufchateau, 75 Virton, 76 Dinant, 77 Namur Arrondissement, 78 Philippeville, 79 Vidin, 80 Vratsa, 81 Montana, 82 Veliko Turnova, 83 Gabrovo, 84 Lovech, 85 Pleven, 86 Ruse, 87 Varna, 88 Dobrich, 89 Razgrad, 90 Silistra, 91 Targovishte, 92 Shumen, 93 Sofia Stolitsa, 94 Blagoevgrad, 95 Kyustendil, 96 Pernik, 97 Sofia, 98 Kurdjali, 99 Pazardzhik, 100 Plovdiv, 101 Smolyan, 102 Stara Zagora, 103 Haskovo, 104 Burgas, 105 Sliven, 106 Yambol, 107 Vaud, 108 Valais, 109 Genève, 110 Bern, 111 Freiburg, 112 Solothurn, 113 Neuchâtel, 114 Jura, 115 Basel Stadt, 116 Basel Landschaft, 117 Aargau, 118 Zürich, 119 Glarus, 120 Schaffhausen, 121 Appenzell-Ausserrhoden, 122 Appenzell-Innerrhoden, 123 St.Gallen, 124 Graubünden, 125 Thurgau, 126 Luzern, 127 Uri, 128 Schwyz, 129 Obwalden, 130 Nidwalden, 131 Zug, 132 Ticino, 133 Cyprus, 134 Praha, 135 Stredocesky, 136 Ceskobudejovicky, 137 Plzensky, 138 Karlovarsky, 139 Ustecky, 140 Liberecky, 141 Kralovehradecky, 142 Pardubicky, 143 Jihlavsky, 144 Brnensky, 145 Olomoucky, 146 Zlinsky, 147 Ostravsky, 148 Stuttgart, Stadtkreis, 149 Böblingen, 150 Esslingen, 151 Göppingen, 152 Ludwigsburg, 153 Rems-Murr-Kreis, 154 Heilbronn, Stadtkreis, 155 Heilbronn, Landkreis, 156 Hohenlohekreis, 157 Schwäbisch Hall, 158 Main-Tauber-Kreis, 159 Heidenheim, 160 Ostalbkreis, 161 Baden-Baden, Stadtkreis, 162 Karlsruhe, Stadtkreis, 163 Karlsruhe, Landkreis, 164 Rastatt, 165 Heidelberg, Stadtkreis, 166 Mannheim, Stadtkreis, 167 Neckar-Odenwald-Kreis, 168 Rhein-Neckar-Kreis, 169 Pforzheim, Stadtkreis, 170 Calw, 171 Enzkreis, 172 Freudenstadt, 173 Freiburg im Breisgau, Stadt, 174 Breisgau-Hochschwarzwald, 175 Emmendingen, 176 Ortenaukreis, 177 Rottweil, 178 Schwarzwald-Baar-Kreis, 179 Tuttlingen, 180 Konstanz, 181 Lörrach, 182 Waldshut, 183 Reutlingen, 184 Tübingen, Landkreis, 185 Zollernalbkreis, 186 Ulm, Stadtkreis, 187 Alb-Donau-Kreis, 188 Biberach, 189 Bodenseekreis, 190 Ravensburg, 191 Sigmaringen, 192 Ingolstadt, Krfr. Stadt, 193 München, Krfr. Stadt, 194 Rosenheim, Krfr. Stadt, 195 Altötting, 196 Berchtesgadener Land, 197 Bad Tölz-Wolfratshausen, 198 Dachau, 199 Ebersberg, 200 Eichstätt, 201 Erding, 202 Freising, 203 Fürstenfeldbruck, 204 Garmisch-Partenkirchen, 205 Landsberg a. Lech, 206 Miesbach, 207 Mühldorf a. Inn, 208 München, Landkreis, 209 Neuburg-Schrobenhausen, 210 Pfaffenhofen an der Ilm, 211 Rosenheim, Landkreis, 212 Starnberg, 213 Traunstein, 214 Weilheim-Schongau, 215 Landshut, Krfr. Stadt, 216 Passau, Krfr. Stadt, 217 Straubing, Krfr. Stadt, 218 Deggen Dorf, 219 Freyung-Grafenau, 220 Kelheim, 221 Landshut, Landkreis, 222 Passau, Landkreis, 223 Reggen, 224 Rottal-Inn, 225 Straubing-Bogen, 226 Dingolfing-Landau, 227 Amberg, Krfr. Stadt, 228

Regensburg, Krfr. Stadt, 229 Weiden i. d. Opf., Krfr. Stadt, 230 Amberg-Sulzbach, 231 Cham, 232 Neumarkt in der Oberpfalz, 233 Neustadt an der Waldnaab, 234 Regensburg, Landkreis, 235 Schwandorf, 236 Tirschenreuth, 237 Bamberg, Krfr. Stadt, 238 Bayreuth, Krfr. Stadt, 239 Coburg, Krfr. Stadt, 240 Hof, Krfr. Stadt, 241 Bamberg, Landkreis, 242 Bayreuth, Landkreis, 243 Coburg, Landkreis, 244 Forchheim, 245 Hof, Landkreis, 246 Kronach, 247 Kulmbach, 248 Lichtentfels, 249 Wunsiedel i. Fichtelgebirge, 250 Ansbach, Krfr. Stadt, 251 Erlangen, Krfr. Stadt, 252 Fürth, Krfr. Stadt, 253 Nürnberg, Krfr. Stadt, 254 Schwabach, Krfr. Stadt, 255 Ansbach, Landkreis, 256 Erlangen-Höchstadt, 257 Fürth, Landkreis, 258 Nürnberger Land, 259 Neustadt an der Aisch-Bad, 260 Roth, 261 Weissenburg-Gunzenhausen, 262 Aschaffenburg, Krfr. Stadt, 263 Schweinfurt, Krfr. Stadt, 264 Würzburg, Krfr. Stadt, 265 Aschaffenburg, Landkreis, 266 Bad Kissingen, 267 Rhön-Grabfeld, 268 Hassberge, 269 Kitzingen, 270 Miltenberg, 271 Main-Spessart, 272 Schweinfurt, Landkreis, 273 Würzburg, Landkreis, 274 Augsburg, Krfr. Stadt, 275 Kaufbeuren, Krfr. Stadt, 276 Kempten (Allgäu), Krfr. Stadt, 277 Memmingen, Krfr. Stadt, 278 Aichach-Friedberg, 279 Augsburg, Landkreis, 280 Dillingen an der Donau, 281 Günzburg, 282 Neu-Ulm, 283 Lindau (Bodensee), 284 Ostallgäu, 285 Unterallgäu, 286 Donau-Ries, 287 Oberallgäu, 288 Berlin-West, Stadt, 289 Berlin-Ost, Stadt, 290 Brandenburg an der Havel, 291 Cottbus, Krfr. Stadt, 292 Frankfurt (Oder), Krfr. Stadt, 293 Potsdam, Krfr. Stadt, 294 Barnim, 295 Dahme-Spreewald, 296 Elbe-Elster, 297 Havelland, 298 Märkisch-Oderland, 299 Oberhavel, 300 Oberspreewald-Lausitz, 301 Oder-Spree, 302 Ostprignitz-Ruppin, 303 Potsdam-Mittelmark, 304 Prignitz, 305 Spree-Neisse, 306 Teltow-Fläming, 307 Uckermark, 308 Bremen, 309 Bremerhaven, Krfr. Stadt, 310 Hamburg, Freie-und Hansestadt, 311 Darmstadt, Krfr. Stadt, 312 Frankfurt am Main, Krfr. Stadt, 313 Offenbach am Main, Krfr. Stadt, 314 Wiesbaden, Krfr. Stadt, 315 Bergstrasse, 316 Darmstadt-Dieburg, 317 Gross-Gerau, 318 Hochtaunuskreis, 319 Main-Kinzig-Kreis, 320 Main-Taunus-Kreis, 321 Odenwaldkreis, 322 Offenbach, Landkreis, 323 Rheingau-Taunus-Kreis, 324 Wetteraukreis, 325 Giessen, Landkreis, 326 Lahn-Dill-Kreis, 327 Limburg-Weilburg, 328 Marburg-Biedenkopf, 329 Vogelsbergkreis, 330 Kassel, Krfr. Stadt, 331 Fulda, 332 Hersfeld-Rotenburg, 333 Kassel, Landkreis, 334 Schwalm-Eder-Kreis, 335 Waldeck-Frankenberg, 336 Werra-Meißner-Kreis, 337 Greifswald, Krfr. Stadt, 338 Neubrandenburg, Krfr. Stadt, 339 Rostock, Krfr. Stadt, 340 Schwerin, Krfr. Stadt, 341 Stralsund, Krfr. Stadt, 342 Wismar, Krfr. Stadt, 343 Bad Doberan, 344 Demmin, 345 Güstrow, 346 Ludwigslust, 347 Mecklenburg-Strelitz, 348 Müritzkreis, 349 Nordvorpommern, 350 Nordwestmecklenburg, 351 Ostvorpommern, 352 Parchim, 353 Rügen, 354 Ücker-Randow, 355 Braunschweig, Krfr. Stadt, 356 Salzgitter, Krfr. Stadt, 357 Wolfsburg, Krfr. Stadt, 358 Gifhorn, 359 Göttingen, 360 Goslar, 361 Helmstedt, 362 Northeim, 363 Osterode am Harz, 364 Peine, 365 Wolfenbüttel, 366 Hannover, Krfr. Stadt, 367 Diepholz, 368 Hameln-Pyrmont, 369 Hannover, Landkreis, 370 Hildesheim, 371 Holzminden, 372 Nienburg (Weser), 373 Schaumburg, 374 Celle, 375 Cuxhaven, 376 Harburg, 377 Lüchow-Dannenberg, 378 Lüneburg, Landkreis, 379 Osterholz, 380 Rotenburg (Wümme), 381 Soltau-Fallingb., 382 Stade, 383 Uelzen, 384 Verden, 385 Delmenhorst, Krfr. Stadt, 386 Emden, Krfr. Stadt, 387 Oldenburg (Oldenburg), Krfr. Stadt, 388 Osnabrück, Krfr. Stadt, 389 Wilhelmshaven, Krfr. Stadt, 390 Ammerland, 391 Aurich, 392 Cloppenburg, 393 Emsland, 394 Friesland, 395 Grafschaft Bentheim, 396 Leer, 397 Oldenburg, Landkreis, 398 Osnabrück, Landkreis, 399 Vechta, 400 Wesermarsch, 401 Wittmund, 402 Düsseldorf, Krfr. Stadt, 403 Duisburg, Krfr. Stadt, 404 Essen, Krfr. Stadt, 405 Krefeld, Krfr. Stadt, 406 Mönchengladbach, Krfr. Stadt, 407 Mülheim a.d.Ruhr, Krfr. Stadt, 408 Oberhausen, Krfr. Stadt, 409 Remscheid, Krfr. Stadt, 410 Solingen, Krfr. Stadt, 411 Wuppertal, Krfr. Stadt, 412 Kleve, 413 Mettmann, 414 Neuss, 415 Viersen, 416 Wesel, 417 Aachen, Krfr. Stadt, 418 Bonn, Krfr. Stadt, 419 Köln, Krfr. Stadt, 420 Leverkusen, Krfr. Stadt, 421 Aachen, Landkreis, 422 Düren, 423 Erftkreis, 424 Euskirchen, 425 Heinsberg, 426 Oberbergischer Kreis, 427 Rheinisch-Bergischer-Kreis, 428 Rhein-Sieg-Kreis, 429 Bottrop, Krfr. Stadt, 430 Gelsenkirchen, Krfr. Stadt, 431 Münster, Krfr. Stadt, 432 Borken, 433 Coesfeld, 434 Recklinghausen, 435 Steinfurt, 436 Warendorf, 437 Bielefeld, Krfr. Stadt, 438 Gütersloh, 439 Herford, 440 Höxter, 441 Lippe, 442 Minden-Lübbecke, 443 Paderborn, 444 Bochum, Krfr. Stadt, 445 Dortmund, Krfr. Stadt, 446 Hagen, Krfr. Stadt, 447 Hamm, Krfr. Stadt, 448 Herne, Krfr. Stadt, 449 Ennepe-Ruhr-Kreis, 450 Hochsauerlandkreis, 451 Märkischer Kreis, 452 Olpe, 453 Siegen-Wittgenstein, 454 Soest, 455 Unna, 456 Koblenz, Krfr. Stadt, 457 Ahrweiler, 458 Altenkirchen (Westerwald), 459 Bad Kreuznach, 460 Birkenfeld, 461 Cochem-Zell, 462 Mayen-Koblenz, 463 Neuwied, 464 Rhein-Hunsrück-Kreis, 465 Rhein-Lahn-

Kreis, 466 Westerwaldkreis, 467 Trier, Krfr. Stadt, 468 Bernkastel-Wittlich, 469 Bitburg-Prüm, 470 Daun, 471 Trier-Saarburg, 472 Frankenthal(Pfalz), Krfr. Stadt, 473 Kaiserslautern, Krfr. Stadt, 474 Landau in der Pfalz, Krfr. Stadt, 475 Ludwigshafen am Rhein, Krfr. Stadt, 476 Mainz, Krfr. Stadt, 477 Neustadt an der Weinstraße, 478 Pirmasens, Krfr. Stadt, 479 Speyer, Krfr. Stadt, 480 Worms, Krfr. Stadt, 481 Zweibrücken, Krfr. Stadt, 482 Alzey-Worms, 483 Bad Dürkheim, 484 Donnersbergkreis, 485 Germersheim, 486 Kaiserslautern, Landkreis, 487 Kusel, 488 Südliche Weinstraße, 489 Ludwigshafen, Landkreis, 490 Mainz-Bingen, 491 Südwestpfalz, 492 Stadtverband Saarbrücken, 493 Merzig-Wadern, 494 Neunkirchen, 495 Saarlouis, 496 Saarpfalz-Kreis, 497 Sankt Wendel, 498 Chemnitz, Krfr. Stadt, 499 Plauen, Krfr. Stadt, 500 Zwickau, Krfr. Stadt, 501 Annaberg, 502 Chemnitzer Land, 503 Freiberg, 504 Vogtlandkreis, 505 Mittlerer Erzgebirgkreis, 506 Mittweida, 507 Stollberg, 508 Aue-Schwarzenberg, 509 Zwickauer Land, 510 Dresden, Krfr. Stadt, 511 Görlitz, Krfr. Stadt, 512 Hoyerswerda, Krfr. Stadt, 513 Bautzen, 514 Meissen, 515 Niederschlesischer Oberlausitzkreis, 516 Riesa-Grossenhain, 517 Löbau-Zittau, 518 Sächsische Schweiz, 519 Weisseritzkreis, 520 Kamenz, 521 Leipzig, Krfr. Stadt, 522 Delitzsch, 523 Döbeln, 524 Leipziger Land, 525 Muldentalkreis, 526 Torgau-Oschatz, 527 Dessau, 528 Anhalt-Zerbst, 529 Bernburg, 530 Bitterfeld, 531 Köthen, 532 Wittenberg, 533 Halle/Saale, Stadtkreis, 534 Burgenlandkreis, 535 Mansfelder Land, 536 Merseburg-Querfurt, 537 Saalkreis, 538 Sangerhausen, 539 Weissenfels, 540 Magdeburg, Krfr. Stadt, 541 Aschersleben-Stassfurt, 542 Bördekreis, 543 Halberstadt, 544 Jerichower Land, 545 Ohrekreis, 546 Stendal, 547 Quedlinburg, 548 Schönebeck, 549 Wernigerode, 550 Altmarkkreis Salzwedel, 551 Flensburg, Krfr. Stadt, 552 Kiel, Krfr. Stadt, 553 Lübeck, Krfr. Stadt, 554 Neumünster, Krfr. Stadt, 555 Dithmarschen, 556 Herzogtum Lauenburg, 557 Nordfriesland, 558 Ostholstein, 559 Pinneberg, 560 Plön, 561 Rendsburg-Eckernförde, 562 Schleswig-Flensburg, 563 Segeberg, 564 Steinburg, 565 Stormarn, 566 Erfurt, 567 Gera, Krfr. Stadt, 568 Jena, Krfr. Stadt, 569 Suhl, Krfr. Stadt, 570 Weimar, Krfr. Stadt, 571 Eichsfeld, 572 Nordhausen, 573 Unstrut-Hainich-Kreis, 574 Kyffhäuserkreis, 575 Schmalkalden-Meiningen, 576 Gotha, 577 Sömmerda, 578 Hildburghausen, 579 Ilm-Kreis, 580 Weimarer Land, 581 Sonneberg, 582 Saalfeld-Rudolstadt, 583 Saale-Holzland-Kreis, 584 Saale-Orla-Kreis, 585 Greiz, 586 Altenburger Land, 587 Eisenach, Krfr. Stadt, 588 Wartburgkreis, 589 København Og Frederiksberg, 590 Københavns Amt, 591 Frederiksborg Amt, 592 Roskilde Amt, 593 Vestsjællands Amt, 594 Storstrøms Amt, 595 Bornholms Amt, 596 Fyns Amt, 597 Sønderjyllands Amt, 598 Ribe Amt, 599 Vejle Amt, 600 Ringkøbing Amt, 601 Aarhus Amt, 602 Viborg Amt, 603 Nordjyllands Amt, 604 Pohja-Eesti, 605 Kesk-Eesti, 606 Kirde-Eesti, 607 Laeaene-Eesti, 608 Louna-Eesti, 609 La Coruna, 610 Lugo, 611 Orense, 612 Pontevedra, 613 Principado de Asturias, 614 Cantabria, 615 Alava, 616 Guipuzcoa, 617 Vizcaya, 618 Comunidad Foral de Navarra, 619 La Rioja, 620 Huesca, 621 Teruel, 622 Zaragoza, 623 Comunidad de Madrid, 624 Avila, 625 Burgos, 626 Leon, 627 Palencia, 628 Salamanca, 629 Segovia, 630 Soria, 631 Valladolid, 632 Zamora, 633 Albacete, 634 Ciudad Real, 635 Cuenca, 636 Guadalajara, 637 Toledo, 638 Badajoz, 639 Cáceres, 640 Barcelona, 641 Girona, 642 Lleida, 643 Tarragona, 644 Alicante, 645 Castellón de la Plana, 646 Valencia, 647 Islas Baleares, 648 Almería, 649 Cádiz, 650 Córdoba, 651 Granada, 652 Huelva, 653 Jaén, 654 Málaga, 655 Sevilla, 656 Región de Murcia, 657 Ceuta, 658 Melilla, 659 Etelae-Savo, 660 Pohjois-Savo, 661 Pohjois-Karjala, 662 Kainuu, 663 Keski-Suomi, 664 Etelae-Pohjanmaa, 665 Pohjanmaa, 666 Keski-Pohjanmaa, 667 Pohjois-Pohjanmaa, 668 Lappi, 669 Uusimaa, 670 Itä-Uusimaa, 671 Varsinais-Suomi, 672 Satakunta, 673 Kanta-Häme, 674 Pirkanmaa, 675 Päijät-Häme, 676 Kymenlaakso, 677 Etelä-Karjala, 678 Åland, 679 Paris, 680 Seine et Marne, 681 Yvelines, 682 Essonne, 683 Hauts de Seine, 684 Seine Saint-Denis, 685 Val de Marne, 686 Val d'Oise, 687 Ardennes, 688 Aube, 689 Marne, 690 Haute Marne, 691 Aisne, 692 Oise, 693 Somme, 694 Eure, 695 Seine Maritime, 696 Cher, 697 Eure et Loir, 698 Indre, 699 Indre et Loire, 700 Loire et Cher, 701 Loiret, 702 Calvados, 703 Manche, 704 Orne, 705 Côte d'Or, 706 Nièvre, 707 Saône et Loire, 708 Yonne, 709 Nord, 710 Pas de Calais, 711 Meurthe et Moselle, 712 Meuse, 713 Moselle, 714 Vosges, 715 Bas Rhin, 716 Haut Rhin, 717 Doubs, 718 Jura, 719 Haute-Saône, 720 Territoire de Belfort, 721 Loire Atlantique, 722 Maine et Loire, 723 Mayenne, 724 Sarthe, 725 Vendée, 726 Côtes d'Armor, 727 Finistère, 728 Ille et Vilaine, 729 Morbihan, 730 Charente, 731 Charente Maritime, 732 Deux Sèvres, 733 Vienne, 734 Dordogne, 735 Gironde, 736 Landes, 737 Lot et Garonne, 738 Pyrénées Atlantiques, 739 Ariège, 740 Aveyron, 741 Haute Garonne, 742 Gers, 743 Lot, 744 Hautes Pyrénées, 745 Tarn, 746 Tarn et Garonne, 747 Corrèze, 748 Creuse, 749 Haute Vienne, 750 Ain, 751 Ardèche, 752 Drôme, 753 Isère, 754 Loire, 755 Rhône, 756 Savoie,

757 Haute Savoie, 758 Allier, 759 Cantal, 760 Haute Loire, 761 Puy de Dome, 762 Aude, 763 Gard, 764 Hérault, 765 Lozère, 766 Pyrénées Orientales, 767 Alpes de Haute Provence, 768 Hautes Alpes, 769 Alpes Maritimes, 770 Bouches du Rhône, 771 Var, 772 Vaucluse, 773 Corse du Sud, 774 Haute Corse, 775 Evros, 776 Xanthi, 777 Rodopi, 778 Drama, 779 Kavala, 780 Imathia, 781 Thessaloniki, 782 Kilkis, 783 Pella, 784 Pieria, 785 Serres, 786 Chalkidiki, 787 Grevena, 788 Kastoria, 789 Kozani, 790 Florina, 791 Karditsa, 792 Larisa, 793 Magnisia, 794 Trikala, 795 Arta, 796 Thesprotia, 797 Ioannina, 798 Preveza, 799 Zakynthos, 800 Kerkyra, 801 Kefallinia, 802 Lefkada, 803 Aitolokarnania, 804 Achaia, 805 Ileia, 806 Voiotia, 807 Evvoia, 808 Evrytania, 809 Fthoitida, 810 Fokida, 811 Argolida, 812 Arkadia, 813 Korinthia, 814 Lakonia, 815 Messinia, 816 Attiki, 817 Lesvos, 818 Samos, 819 Chios, 820 Dodekanisos, 821 Kyklades, 822 Irakleio, 823 Lasithi, 824 Rethymni, 825 Chania, 826 Budapest, 827 Pest, 828 Fejer, 829 Komarom-Esztergom, 830 Veszprem, 831 Győr-Moson-Sopron, 832 Vas, 833 Zala, 834 Baranya, 835 Somogy, 836 Tolna, 837 Borsod-Abaúj-Zemplén, 838 Heves, 839 Nógrád, 840 Hajdu-Bihar, 841 Jász-Nagykun-Szolnok, 842 Szabolcs-Szatmár-Bereg, 843 Bács-Kiskun, 844 Bekes, 845 Csongrád, 846 Border, 847 Midland, 848 West, 849 Dublin, 850 Mid-East, 851 Mid-West, 852 South-East (Irl), 853 South-West (Irl), 854 Torino, 855 Vercelli, 856 Biella, 857 Verbano-Cusio-Ossola, 858 Novara, 859 Cuneo, 860 Asti, 861 Alessandria, 862 Valle d'Aosta, 863 Imperia, 864 Savona, 865 Genova, 866 La Spezia, 867 Varese, 868 Como, 869 Lecco, 870 Sondrio, 871 Milano, 872 Bergamo, 873 Brescia, 874 Pavia, 875 Lodi, 876 Cremona, 877 Mantova, 878 Bolzano-Bozen, 879 Trento, 880 Verona, 881 Vicenza, 882 Belluno, 883 Treviso, 884 Venezia, 885 Padova, 886 Rovigo, 887 Pordenone, 888 Udine, 889 Gorizia, 890 Trieste, 891 Piacenza, 892 Parma, 893 Reggio Nell'Emilia, 894 Modena, 895 Bologna, 896 Ferrara, 897 Ravenna, 898 Forlì-Cesena, 899 Rimini, 900 Massa-Carrara, 901 Lucca, 902 Pistoia, 903 Firenze, 904 Prato, 905 Livorno, 906 Pisa, 907 Arezzo, 908 Siena, 909 Grosseto, 910 Perugia, 911 Terni, 912 Pesaro e Urbino, 913 Ancona, 914 Macerata, 915 Ascoli Piceno, 916 Viterbo, 917 Rieti, 918 Roma, 919 Latina, 920 Frosinone, 921 L'Aquila, 922 Teramo, 923 Pescara, 924 Chieti, 925 Isernia, 926 Campobasso, 927 Caserta, 928 Benevento, 929 Napoli, 930 Avellino, 931 Salerno, 932 Foggia, 933 Bari, 934 Taranto, 935 Brindisi, 936 Lecce, 937 Potenza, 938 Matera, 939 Cosenza, 940 Crotone, 941 Catanzaro, 942 Vibo Valentia, 943 Reggio di Calabria, 944 Trapani, 945 Palermo, 946 Messina, 947 Agrigento, 948 Caltanissetta, 949 Enna, 950 Catania, 951 Ragusa, 952 Siracusa, 953 Sassari, 954 Nuoro, 955 Oristano, 956 Cagliari, 957 Alytaus, 958 Kauno, 959 Klaipėdos, 960 Marijampolės, 961 Panevezio, 962 Siauliai, 963 Taurages, 964 Telsiai, 965 Utenos, 966 Vilniaus, 967 Luxembourg, 968 Riga, 969 Vidzeme, 970 Kurzeme, 971 Zemgale, 972 Latgale, 973 Malta, 974 Gozo and Comino, 975 Oost-Groningen, 976 Delfzijl en Omgeving, 977 Overig Groningen, 978 Noord-Friesland, 979 Zuidwest-Friesland, 980 Zuidoost-Friesland, 981 Noord-Drenthe, 982 Zuidoost-Drenthe, 983 Zuidwest-Drenthe, 984 Noord-Overijssel, 985 Zuidwest-Overijssel, 986 Twente, 987 Veluwe, 988 Achterhoek, 989 Arnhem/Nijmegen, 990 Zuidwest-Gelderland, 991 Flevoland, 992 Utrecht, 993 Kop van Noord-Holland, 994 Alkmaar en Omgeving, 995 IJmond, 996 Agglomeratie Haarlem, 997 Zaanstreek, 998 Groot-Amsterdam, 999 Het Gooi en Vechtstreek, 1000 Aggl. Leiden en Bollenstr, 1001 Agglomeratie S-Gravenhage, 1002 Delft en Westland, 1003 Oost Zuid-Holland, 1004 Groot-Rijnmond, 1005 Zuidoost Zuid-Holland, 1006 Zeeuwsch-Vlaanderen, 1007 Overig Zeeland, 1008 West-Noord-Brabant, 1009 Midden-Noord-Brabant, 1010 Noordoost-Noord-Brabant, 1011 Zuidoost-Noord-Brabant, 1012 Noord-Limburg, 1013 Midden-Limburg, 1014 Zuid-Limburg, 1015 Oslo, 1016 Åkershus, 1017 Hedmark, 1018 Oppland, 1019 Østfold, 1020 Buskerud, 1021 Vestfold, 1022 Telemark, 1023 Aust-Agder, 1024 Vest-Agder, 1025 Rogaland, 1026 Hordaland, 1027 Sogn Og Fjordane, 1028 Møre Og Romsdal, 1029 Sør-Trøndelag, 1030 Nord-Trøndelag, 1031 Nordland, 1032 Troms, 1033 Finnmark, 1034 Jeleniogarsko-Walbrzyski, 1035 Legnicki, 1036 Wrocławski, 1037 M. Wrocław, 1038 Bydgoski, 1039 Toruńsko-Wrocławski, 1040 Białkopodlaski, 1041 Chełmsko-Zamojski, 1042 Lubelski, 1043 Gorzowski, 1044 Zielonogarski, 1045 Łódzki, 1046 Piotrkowski-Skierniewicki, 1047 M. Łódz, 1048 Krakowski-Tarnowski, 1049 Nowosadecki, 1050 M. Kraków, 1051 Ciechanowski-Płocki, 1052 Ostrołęcki-Siedlecki, 1053 Warszawski, 1054 Radomski, 1055 M. Warszawa, 1056 Opolski, 1057 Rzeszowski-Tarnobrzeczki, 1058 Krasnienski-Przemyski, 1059 Białostocko-Suwalski, 1060 Łomżyński, 1061 Słupski, 1062 Gdąński, 1063 Gdańsk-Gdynia-Sopot, 1064 Palnośnoslaski, 1065 Południowoslaski, 1066 Centralny Slaski, 1067 Świetokrzyski, 1068 Elblaski, 1069 Olsztynski, 1070 Elcki, 1071 Piłski, 1072 Poznański, 1073 Kaliski, 1074 Koninski, 1075 M. Poznań, 1076 Szczeciński, 1077 Koszaliński, 1078 Minhó-Lima, 1079 Cavado, 1080

Ave, 1081 Grande Porto, 1082 Tamega, 1083 Entre Douro e Vouga, 1084 Douro, 1085 Alto Tras-Os-Montes, 1086 Baixo Vouga, 1087 Baixo Mondego, 1088 Pinhal Litoral, 1089 Pinhal Interior Norte, 1090 Dao-Lafões, 1091 Pinhal Interior Sul, 1092 Serra da Estrela, 1093 Beira Interior Norte, 1094 Beira Interior Sul, 1095 Cova Da Beira, 1096 Oeste, 1097 Grande Lisboa, 1098 Peninsula de Setúbal, 1099 Médio Tejo, 1100 Lezíria do Tejo, 1101 Alentejo Litoral, 1102 Alto Alentejo, 1103 Alentejo Central, 1104 Baixo Alentejo, 1105 Algarve, 1106 Bacau, 1107 Botosani, 1108 Iasi, 1109 Neamt, 1110 Suceava, 1111 Vaslui, 1112 Braila, 1113 Buzau, 1114 Constanta, 1115 Galati, 1116 Tulcea, 1117 Vrancea, 1118 Arges, 1119 Calarasi, 1120 Dambovita, 1121 Giurgiu, 1122 Ialomita, 1123 Prahova, 1124 Teleorman, 1125 Dolj, 1126 Gorj, 1127 Mehedinti, 1128 Olt, 1129 Valcea, 1130 Arad, 1131 Caras-Severin, 1132 Hunedoara, 1133 Timis, 1134 Bihor, 1135 Bistrita-Nasaud, 1136 Cluj, 1137 Maramures, 1138 Satu Mare, 1139 Salaj, 1140 Alba, 1141 Brasov, 1142 Covasna, 1143 Harghita, 1144 Mures, 1145 Sibiu, 1146 Bucuresti, 1147 Ilfov, 1148 Stockholms Län, 1149 Uppsala Län, 1150 Södermanlands Län, 1151 Östergötlands Län, 1152 Örebro Län, 1153 Västmanlands Län, 1154 Blekinge Län, 1155 Skane Län, 1156 Värmlands Län, 1157 Dalarnas Län, 1158 Gävleborgs Län, 1159 Västernorrlands Län, 1160 Jämtlands Län, 1161 Västerbottens Län, 1162 Norrbottens Län, 1163 Jönköpings Län, 1164 Kronobergs Län, 1165 Kalmar Län, 1166 Gotlands Län, 1167 Hallands Län, 1168 Västra Götalands Län, 1169 Pomurska, 1170 Podravska, 1171 Koroska, 1172 Savinjska, 1173 Zasavska, 1174 Spodnje-posavska, 1175 Dolenjska, 1176 Osrednjeslovenska, 1177 Gorenjska, 1178 Notranjsko-Kraska, 1179 Goriska, 1180 Obalno-Kraska, 1181 Bratislavsky Kraj, 1182 Tmavsky Kraj, 1183 Trenciansky Kraj, 1184 Nitriansky Kraj, 1185 Zilinsky Kraj, 1186 Banskobystricky Kraj, 1187 Presovsky Kraj, 1188 Kosicky Kraj, 1189 Hartlepool and Stockton-On-Tees, 1190 South Teesside, 1191 Darlington, 1192 Durham, 1193 Northumberland, 1194 Tyneside, 1195 Sunderland, 1196 West Cumbria, 1197 East Cumbria, 1198 Halton and Warrington, 1199 Cheshire, 1200 Greater Manchester South, 1201 Greater Manchester North, 1202 Blackburn with Darwen, 1203 Blackpool, 1204 Lancashire, 1205 East Merseyside, 1206 Liverpool, 1207 Sefton, 1208 Wirral, 1209 Kingston upon Hull, 1210 East Riding of Yorkshire, 1211 North and North-East Lincolnshire, 1212 York, 1213 North Yorkshire, 1214 Barnsley, Doncaster, Roth, 1215 Sheffield, 1216 Bradford, 1217 Leeds, 1218 Calderdale, Kirklees, Wakefield, 1219 Derby, 1220 East Derbyshire, 1221 South and West Derbyshire, 1222 Nottingham, 1223 North Nottinghamshire, 1224 South Nottinghamshire, 1225 Leicester, 1226 Leicestershire, Rutlan, 1227 Northamptonshire, 1228 Lincolnshire, 1229 Herefordshire, 1230 Worcestershire, 1231 Warwickshire, 1232 Telford and Wrekin, 1233 Shropshire, 1234 Stoke-On-Trent, 1235 Staffordshire, 1236 Birmingham, 1237 Solihull, 1238 Coventry, 1239 Dudley and Sandwell, 1240 Walsall and Wolverhampton, 1241 Peterborough, 1242 Cambridgeshire, 1243 Norfolk, 1244 Suffolk, 1245 Luton, 1246 Bedfordshire, 1247 Hertfordshire, 1248 Southend-On-Sea, 1249 Thurrock, 1250 Essex, 1251 Inner London-West, 1252 Inner London-East, 1253 Outer London-East and North East, 1254 Outer London-South, 1255 Outer London-West and North West, 1256 Berkshire, 1257 Milton Keynes, 1258 Buckinghamshire, 1259 Oxfordshire, 1260 Brighton and Hove, 1261 East Sussex, 1262 Surrey, 1263 West Sussex, 1264 Portsmouth, 1265 Southampton, 1266 Hampshire, 1267 Isle of Wight, 1268 Medway, 1269 Kent, 1270 Bristol, 1271 North and North-East Somerset, South, 1272 Gloucestershire, 1273 Swindon, 1274 Wiltshire, 1275 Bournemouth and Poole, 1276 Dorset, 1277 Somerset, 1278 Cornwall, Isle Of Scilly, 1279 Plymouth, 1280 Torbay, 1281 Devon, 1282 Isle of Anglesey, 1283 Gwynedd, 1284 Conwy and Denbighshire, 1285 South West Wales, 1286 Central Valleys, 1287 Gwent Valleys, 1288 Bridgend, Neath Port Talb, 1289 Swansea, 1290 Monmouthshire, Newport, 1291 Cardiff, Vale Of Glamorgan, 1292 Flintshire and Wrexham, 1293 Powys, 1294 Aberdeenshire, North East, 1295 Angus, Dundee City, 1296 Clackmannanshire and Fife, 1297 East Lothian and Midlothian, 1298 Scottish Borders, 1299 Edinburgh, 1300 Falkirk, 1301 Perth, Kinross, Stirling, 1302 West Lothian, 1303 East and West Dunbartonshire, 1304 Dumfries and Galloway, 1305 East and North Ayrshire, Mainland, 1306 Glasgow, 1307 Inverclyde, East Renfrews, 1308 North Lanarkshire, 1309 South Ayrshire, 1310 South Lanarkshire, 1311 Caithness, Sutherland, Ross, 1312 Badenoch, Strathspey, Loc, 1313 Lochaber, Skye, Lochalsh, Ar, 1314 Eilean Siar (Western Isle), 1315 Orkney Islands, 1316 Shetland Islands, 1317 Belfast, 1318 Outer Belfast, 1319 East of Northern Ireland, 1320 North of Northern Ireland, 1321 West and South of Northern Ireland, 1322 Shqipëria, 1323 Bosnia I Hercegovina, 1324 Minsk, 1325 Witebsk, 1326 Mogiljow, 1327 Gomel, 1328 Brest, 1329 Grodno, 1330 Zagreb, 1331 Dalmacija, 1332 Island, 1333 Liechtenstein, 1334 Moldova, 1335 Makedonija,

1336 Archangelskaja Oblast, 1337 Vologodskaja Oblast, 1338 Murmanskaja Oblast, 1339 Karelijal, Respublika, 1340 Komi, Respublika, 1341 Neneckij Avtonomnyi Okrug, 1342 Leningradskaja Oblast, 1343 Sankt-Peterburg, Gorod, 1344 Novgorodskaja Oblast, 1345 Pskovskaja Oblast, 1346 Brjanskaja Oblast, 1347 Vladimirskaia Oblast, 1348 Ivanovskaja Oblast, 1349 Kaluzskaja Oblast, 1350 Kostromskaja Oblast, 1351 Moskva Oblast, 1352 Moskva, Gorod, 1353 Orlovskaja Oblast, 1354 Rjasan Oblast, 1355 Smolenskaja Oblast, 1356 Tverskaja Oblast, 1357 Tulskaia Oblast, 1358 Jaroslavskaja Oblast, 1359 Belgorodskaja Oblast, 1360 Kurskaja Oblast, 1361 Lipeckaja Oblast, 1362 Kaliningrad, 1363 Other Russia, 1364 Türkiye, 1365 South West, 1366 South, 1367 Donezk-Dnepr, 1368 West, 1369 Serbia, 1370 Voivodina, 1371 Kosovo, 1372 Montenegro, 1373 ROW.

Chapter 1

Introduction

Why do some European regions produce so much more output per capita than others? This question has been extensively debated among regional economists and politicians for decades. Large income differences within Europe are not only visible when comparing agglomeration and periphery – be it at regional, national and international level. Incomes also differ considerably among agglomerated regions, especially across countries. In 2002, the London agglomeration¹ generated a gross domestic product (GDP) of roughly 36,000 PPP-Euro² per capita, whereas the Sofia agglomeration accounted for only 11,000 PPP-Euro per capita.³ When comparing agglomerated to peripheral regions, the gap widens further. In 2002, Smolyan, a Bulgarian region at the border to Greece, achieved a GDP per capita of only 4,900 PPP-Euro.

In the preamble of the Treaty of Rome, signed in 1957, the economic cohesion among member states of the European Union (EU) is emphasized as being an important task for integrating the European Economic Community. Subsequently, European politicians have raised two questions: Are the observed income differences tolerably small or intolerably large? And how could a minimization of income differences across nations and regions be achieved? The fathers of the Treaty of Rome believed in classical economic theories rather than in political intervention and, hence, trusted in market forces to ensure convergence among regions.⁴ Not until its Northern Enlargement in 1973, and thus with widening regional disparities, the

¹The London agglomeration consists of Inner and Outer London regions, according to the *Nomenclature des Unités Territoriales pour la Statistique* (NUTS) by Eurostat (1999).

²GDP is expressed in purchasing power parities (PPP) to assure comparability across countries.

³*Eurostat's New Cronos* database (<http://epp.eurostat.ec.europa.eu>, last visit March 2008).

⁴For instance, in economic trade theory, interaction among economies was per se regarded as welfare enhancing for all trading partners (e.g. Ricardo (1817), Heckscher (1919), and Ohlin (1933)), especially for those engaging in trade integration (Viner (1950)).

European Economic Community established a common regional policy. At the time, financial support was primarily assigned to regions lagging behind. In the following decades, this new European policy became a vast catalogue of regional, structural and cohesion policies. With every new integration round the regional funds grew considerably.⁵ For the period 2007-2013, 36% of the total EU budget is assessed for various policy initiatives which are intended to promote convergence of regional income levels.⁶ But are these expenses justified? Or is the uneven spatial income distribution simply the consequence of a spatial economic equilibrium which is economically efficient and, in sum, welfare enhancing?

To answer these questions, it is essential to understand the economic forces that form the spatial structure of economic activity. Thirty years ago, the introduction of the *New Trade Theory* has shattered the fundament of traditional trade theory. Inter alia, theorists have relaxed the neoclassical assumption of perfect competition and introduced economies of scale into the framework of analysis (Spence (1976), Dixit & Stiglitz (1977), Krugman (1979, 1980)). Almost two decades ago, Krugman (1991) explicitly implemented spatial links into trade models and launched the discipline of the *New Economic Geography* (NEG). These theories convincingly show that not all interacting regions benefit from trade when economies of scale and transport costs are introduced. Regional income differences are seen as the consequence of economic interactions between regions. The key innovation of the *New Trade Theory* is to put a major argument on regional market accessibility when explaining these differences. Stated differently, trade integration reduces trade barriers and transport costs in principle. However, it is the location of a region, e.g. its market proximity, that determines whether a trading partner gains or loses from trade integration. The rationale is as follows: Existing economies of scale favor agglomerations because of short distances between economic agents and a better transport infrastructure. Both yield lower transport costs and, hence, higher income levels in economic centers compared to the periphery.

The *New Trade Theory* has also changed the paradigm of European regional policy. Nowadays, the European Commission puts less emphasis on traditional instruments of regional policy, but stresses the importance of projects improving transport infrastructure links. For example, the Commission prioritized its *Trans-European Networks* (TEN) program and provided a large financial supplement for it. With

⁵For an analysis of political bargaining as major explanatory force behind this trend in European regional policy, see Baldwin *et al.* (1997).

⁶This accounts for 50 billion Euro each year. For details on the allocation of financial aids among member states see http://ec.europa.eu/regional_policy/policy/fonds/index_en.htm (last visit March 2008).

the implementation of the financial perspective for 2007-2013, Brussels started to assign financial support not only to poorer regions, but also to agglomerated regions (see European Commission (2004)). However, there are also sceptical remarks in the economic literature questioning this strategy. For instance, Martin (1998) presumes that improving transportation links strengthens agglomerations even further since the economies of scale facilitate a broader and cheaper distribution of goods. According to this argument, peripheral regions would be even more disadvantaged and European income differences would widen further if transport links between agglomerations and the periphery were ameliorated. Stated differently, although all regions benefit from a better transport infrastructure, it would not help to eliminate the relative regional income differences across Europe.

Since economists and politicians debate on which conclusions to draw from theory, it is important to test whether the major argument of the *New Trade Theory* holds true. Does market proximity really have a positive impact on regional income levels or are the latter rather explained by alternative variables? If we proved the theoretical relationship to be right, we would have found evidence that improving market accessibility for a region in fact should yield higher income levels. However, if other explaining variables or unobservable region-specifics play an even more important role, we would need to conjecture that the described policy measures alone are not suitable to promote higher incomes.

In this monograph, we show how and to which extent regional equilibrium prices – which are a proxy for the degree of market accessibility of a region – affect the European income distribution. We find an answer using an innovative empirical approach which overcomes some of the loose ends in the literature on the empirical validation of the *New Trade Theory*. Most importantly, we unite the state-of-the-art-approaches of Hanson (2005) and Redding & Venables (2004) to a comprehensive framework of analysis. We follow the idea of Knaap (2004) and combine econometric tools with a Computable General Equilibrium (CGE) Model, namely the static *CGEurope Model* developed by Bröcker (1995, 1998b). Our strategy encloses the following four major advantages.

1. We do not need to rely on explicit regional trade and regional price data which are not available in statistics databases for Europe. In our model, both are implicitly given by the model's equilibrium solution. Due to this calibration outcome we can include all model dimensions into our analysis.
2. Our approach allows to calculate border impediments to international trade on the basis of the empirical results. The reason is that our transport costs

implementation consists of two components: distance costs and an ad-valorem tariff-equivalent. Border effects appear to be more relevant than distance costs for the development of trade between regions.⁷ Therefore, this is a strong asset of our approach.

3. We can obtain a fully parameterized model. Our model reflects an economic equilibrium which is calibrated by real world data. This permits to check whether estimated parameter values imply plausible distance and transport cost functions. On the contrary, authors in empirical CGE literature often need to assume certain plausible values for the key parameters in order to be able to calibrate their respective models. This strategy is also followed by some authors who combine simulation and estimation approaches, e.g. Redding & Sturm (2005).
4. In contrast to most existing studies, we do not have to assume real wage equalization across regions in order to come to an empirical specification of the wage equation.

Finally, we use a disaggregated regional data set which covers all the 27 member states of the enlarged European Union. This has not been done yet in literature.

The remainder of this monograph is organized as follows. Chapter 2 surveys related empirical literature. Chapter 3 presents the theoretical CGE model used and derives the empirical strategy. Chapter 4 comprises our econometric approach. We elaborate the advantages and shortcomings of cross-sectional and panel approaches, and discuss how several econometric problems can be addressed. Our data set is presented in chapter 5. Here, we also present solutions to potential problems which might arise when using data gathered without preprocessing. Chapter 6 discusses the results of the econometric analysis. Chapter 7 concludes.

⁷See e.g. Bröcker (1984), McCallum (1995), or Helliwell (1996).

Chapter 2

A Critical Literature Review

This chapter provides a critical review of existing empirical work closely related to our approach. We concentrate on studies that seek to prove a positive impact of accessibility on income levels.¹ The chapter is organized in three main parts. The first part gives a short theoretical introduction to spatial trade models à la Dixit & Stiglitz (1977) in order to present the workhorse of existing empirical studies. In the second part, empirical work itself is briefly reviewed to show what has already been done. In particular, we discuss the results as well as the strategies of researchers to overcome two obstacles. On the one hand, they often need to rely on restrictive assumptions in order to derive empirically tractable equations. On the other hand, they are forced to solve problems of missing data which are mandatory for the estimation. The third section provides a thorough critique of the work presented. Because of the condensed wrap-up, the reader should be aware of the problems and shortcomings of the existent approaches as well as of the possibilities in overcoming them with our suggestion. Finally, the chapter concludes with a summary.

2.1 Theoretical Foundations

Monopolistic competition trade models date back to Chamberlin (1933), Spence (1976) and Dixit & Stiglitz (1977). They were further developed to the *New Trade Theory* (Krugman (1979) and Krugman (1980)) and to the NEG (Krugman (1991)).²

¹A complete review of the literature can be found in Hanson (2001), Brakman *et al.* (2001), Overman *et al.* (2003), or Head & Mayer (2004). Further prominent strands of literature test whether market accessibility attracts factor inflows (e.g. Head & Mayer (2002b), Crozet (2004), or Tabuchi & Thisse (2002)) and whether the production or the demand side of an economy is stimulated by magnification effects on the home market (Davis & Weinstein (1999), Davis & Weinstein (2003), Head & Ries (2001), or Trionfetti (2001)).

²A review of the basic models and a large number of extensions and mutations can be found in Fujita *et al.* (2001) and Baldwin *et al.* (2003)).

The latter are based on intuitive linkages of economic activities. The original idea comprised relations between firms. Hirschman (1958) separates two kinds of linkages between upstream and downstream industries: so-called forward and backward linkages. A backward linkage describes the process of an additional producer settling into a region. This augments the demand for intermediate goods which are provided by the resident producers. Hence, these producers face economies of scale through lower unit costs (which is the so-called home market effect). Forward linkages arise when the described lower unit costs are passed on to the final goods producers who can lower their price and increase their sales accordingly (which is the so-called price index effect). Myrdal (1957) describes a process of cumulative causation that results from the two linkages. The described forward process of decreasing prices and rising sales leads to an increasing demand for intermediate goods. This, in turn, augments backward linkages. From this circular causation mechanism evolves the following hypothesis. Agglomerations should always benefit more from interaction among regions than peripheral regions do because the proximity of economic agents to the markets is bigger. That is why the linkages described are also called centripetal forces since the underlying mechanisms lead to a concentration of economic activity. Beyond these forces, models à la Dixit & Stiglitz (1977) also introduce a centrifugal force which favors a more dispersed spatial allocation of economic activity. This is done via implementing an immobile factor of production that prevents absolute agglomeration via a competition effect.

2.1.1 Monopolistic Competition Trade Models

In the following, we analytically present the basic mechanisms and conclusions of a spatial version of the Dixit & Stiglitz (1977) model.³ The three major equilibrium equations of the model are derived because they serve as the workhorse for later presented empirical studies.

Assumptions. We consider an economy made of multiple regions, each consisting of one production sector. The production sector produces a number of varieties of a differentiated industrial good under economies of scale and monopolistic competition. Goods can be traded among regions. Firms are entirely owned by the households.

³This section largely builds on Fujita *et al.* (2001). Whenever mathematical notation is used in this chapter it differs from the mentioned monograph in order to match the notation in the subsequent chapters.

Demand Side. Initially, we focus on the economy of one single region. Households choose the composition of their consumption bundle C_M following an expenditure minimization problem. According to Dixit & Stiglitz (1977) it consists of h single varieties that make up the variety continuum. The quantity consumed of a respective variety i is denoted by c_i and $0 < \rho < 1$ is the inverse measure of the intensity of the preference for variety. Analytically, we have

$$C_M = \left(\int_0^h c_i^\rho di \right)^{1/\rho}. \quad (2.1)$$

Then, we define q_M as the expenditure level or price index for C_M :

$$q_M = \min_{\tilde{c}_i} \left(\int_0^h p_i \tilde{c}_i di \right), \quad (2.2)$$

where p_i is the price of a variety produced and \tilde{c}_i is the expenditure minimizing consumption of variety i . The optimization restriction is given by $\int_0^h c_i^\rho di = 1$ and assures that a household achieves an utility level of one. Optimizing with respect to \tilde{c}_i , gives the expenditure minimizing consumption of variety i

$$\tilde{c}_i = \left(\frac{p_i}{q_M} \right)^{-\sigma}. \quad (2.3)$$

Parameter $\sigma := 1/(1-\rho) > 1$ is denoting the constant elasticity of substitution. If ρ is close to 1, manufacturing varieties are nearly perfect substitutes. As ρ decreases toward 0, σ is decreasing toward 1 and the desire to consume a greater variety of manufacturing goods increases. According to equation (2.3), a higher price index for manufacturing goods decreases the relative price of a variety and thus yields a higher consumption level per variety. A higher production price p_i has the opposite effect.

Using equation (2.3) in $q_M = \int_0^h p_i \tilde{c}_i di$, we obtain for the price index

$$q_M = \left(\int_0^h p_i^{(1-\sigma)} di \right)^{\frac{1}{1-\sigma}}. \quad (2.4)$$

For uniform prices $p_i = p \forall i$ we can write down the composite price index as

$$q_M = (hp^{1-\sigma})^{\frac{1}{1-\sigma}} = h^{\frac{1}{1-\sigma}} p. \quad (2.5)$$

This term describes the preference for variety of consumers. The higher the number of varieties h , the smaller is expression $h^{\frac{1}{1-\sigma}}$ and thus the price index for manufac-

turing goods.

For later purposes, we need to state the expenditure for variety i . Expanding equation (2.3) by p_i/q_M yields:

$$\frac{p_i \tilde{c}_i}{q_M} = \left(\frac{p_i}{q_M} \right)^{1-\sigma}. \quad (2.6)$$

We define the value of final demand for manufacturing goods as D . Using this in equation (2.6) gives the expenditure per variety,

$$p_i c_i = \left(\frac{p_i}{q_M} \right)^{1-\sigma} D. \quad (2.7)$$

Supply Side. Firms produce under linear costs with fixed labor costs F per variety and a variable marginal labor input requirement l per unit of a variety produced. Factor prices for variable and fixed labor inputs are identical and given by the wage rate w . Free entry drives profits down to zero, thus $wF = x(p(x) - wl)$, where x is the output quantity and $p(x)$ is the price as a function of output.

No firm will choose to produce the same variety supplied by another firm due to economies of scale, consumers' preference for variety, and the unlimited number of potential varieties. Hence, firms maximize profits according to:

$$\max_x \{p(x)x - w(lx + F)\}. \quad (2.8)$$

Solving the maximization problem yields the output price. It is calculated via the first order condition with respect to x . Using $(dp/dx)(x/p) = -1/\sigma$, the price is given by

$$p = \frac{\sigma}{\sigma - 1} wl. \quad (2.9)$$

Thus, prices are given by a constant mark-up on marginal costs. Because profits are equal to zero, we can determine the output quantity produced by

$$x = (\sigma - 1) \frac{F}{l}. \quad (2.10)$$

Given a constant output level $x = \bar{x} \forall i$, the number of varieties produced is given by

$$h = \frac{L}{\sigma F}, \quad (2.11)$$

where L is the labor stock.

Transportation, Trade, and Equilibrium. We now turn to the economic interaction among regions. Transport from producer to consumer is costly. When a good is shipped from region s to region r , the iceberg transport cost à la Samuelson (1952) is applied. The cost is simply modelled by the assumption that a fraction of the good itself melts away during transportation without creating revenues in a transport sector. This melting share is given by $(1 - 1/\tau_{sr})$, where $\tau_{sr} \in [1, \infty)$ describes the transport cost factor. Consumer prices p_{sr} – or cost insurance freight (c.i.f.) including prices – in region r are the free on board (f.o.b.) goods prices p_s for goods purchased in region s multiplied by τ_{sr} .

The equilibrium of the model is determined by the clearing of all markets. It consists of a system of three equations. These are the price index equation, the trade equation, and the wage equation.

A power transformation of the prices for the composite tradable ($q_r^{1-\sigma}$) equals the weighted sum of power transformed c.i.f. demand prices ($(p_s \tau_{sr})^{1-\sigma}$) which are to be paid in region r for manufacturing goods stemming from all other regions s . Analogous to equation (2.4), the corresponding price index can be written as

$$q_r = \left(\sum_s h_s (p_s \tau_{sr})^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \quad (2.12)$$

The price index is small, if only few varieties have to be imported from regions s , and if c.i.f. demand prices in region r are low.

In equilibrium, we need to assure that the value of tradables supply in region s equals the sum of trade flows X_{sr} from region s to all regions including s itself. Using $x_s = \bar{x} \forall r$, we have

$$h_s p_s \bar{x} = \sum_r X_{sr}. \quad (2.13)$$

Using equation (2.7) and replacing q_M by q_r , the gravity trade model is obtained:

$$X_{sr} = h_s \left(\frac{p_s \tau_{sr}}{q_r} \right)^{1-\sigma} D_r. \quad (2.14)$$

Hence, the value of goods imported by a region r from a region s equals the product of the number of varieties produced in regions s , multiplied by the expenditure shares for those varieties in region r , and multiplied by the overall expenditure for manufacturing goods in region r .

The equilibrium condition for supply in region s leads to the wage equation. We

insert the gravity equation (2.14) into equation (2.13), and solve it for p_s :

$$p_s^\sigma = \frac{1}{\bar{x}} \sum_r \left(\frac{\tau_{sr}}{q_r} \right)^{1-\sigma} D_r. \quad (2.15)$$

Then, we use price equation (2.9) in equation (2.15), solve the latter for w , and gather all constants terms in η in order to state the wage equation:

$$w_s = \eta \left[\sum_r \left(\frac{\tau_{sr}}{q_r} \right)^{1-\sigma} D_r \right]^{1/\sigma}. \quad (2.16)$$

Since $\sigma > 1$, regional wages in region s rise with lower transport costs between regions, with a higher price index for the composite tradables from regions r , and with higher expenditures for manufacturing goods in regions r .

Equations (2.12), (2.14), and (2.16) are the key equations of the Dixit-Stiglitz-model and are the underpinning of the empirical literature. The wage equation is especially important because it describes the relationship between regional incomes and market proximity. An empirical test provides an answer to the questions raised in the beginning of this monograph.

2.1.2 New Economic Geography

Departing from a model à la Dixit & Stiglitz (1977), Krugman (1991) showed how interactions among increasing returns at the level of the firm, transport costs, and factor mobility can cause a spatial economic structure to emerge and change. In his model, he therefore assumes that manufacturing workers are mobile between regions. Only in the long run no migration incentive is left. The above described firm related linkages now evolve from interaction between firms and households.

According to Neary (2001), the linkages work as follows. The more firms produce in a single region, the bigger the number of varieties of local goods and the bigger the consumption space for households. Since local goods are not subject to transportation costs a household's real income increases. Higher real incomes attract new consumers and workers to a region. This is NEG's forward linkage or price index effect. Newly attracted consumers make local firms face a higher demand who can then save on production costs because of economies of scale (backward linkage). The latter mechanism is also known as NEG's home market effect. The magnitudes of the described effects are determined, for instance, by the level of transport costs. Note that these linkages are exclusively of a pecuniary nature. First nature effects

(such as spatial geographic heterogeneity) and technological externalities (such as knowledge spillover) are not modelled explicitly.

By the introduction of a second (traditional) production sector which produces goods under perfect competition, Krugman introduces a dispersive force into the model. Households spend a fraction ε of their income Y on the manufacturing goods and a fraction $(1 - \varepsilon)$ on the traditional goods. Workers in the traditional sector are assumed to be immobile and to be paid an identical wage across regions. Since this fraction of consumers never relocates, manufacturing firms in an agglomerated region face higher competition on the selling market compared to the periphery. This, by trend, yields lower nominal wages for their manufacturing workers. The so-called competition effect weakens the centripetal forces mentioned above.

The strong assumptions in the Krugman-model lead to two alternative long run equilibrium solutions. The first is characterized by equilibrated real wages. This means that in the long run workers have no incentive to move between regions and regional incomes are dispersed. The second equilibrium solution is characterized by catastrophic agglomeration, thus leaving behind regions which are entirely depleted, i.e. where no modern production takes place. The two solutions can be described by the following relations:

$$\frac{w_r}{q_r^\varepsilon} \leq \bar{\omega} \quad ; \quad h_r \geq 0 \quad ; \quad h_r \left(\bar{\omega} - \frac{w_r}{q_r^\varepsilon} \right) = 0, \quad (2.17)$$

where $\bar{\omega}$ denotes the equilibrium real wage. If in equilibrium the real wages in one region are lower than in other regions ($\frac{w_r}{q_r^\varepsilon} < \bar{\omega}$), no manufacturing varieties are produced in this region ($h_r = 0$). The equilibrium condition $h_r \left(\bar{\omega} - \frac{w_r}{q_r^\varepsilon} \right) = 0$ is fulfilled. If $\frac{w_r}{q_r^\varepsilon} = \bar{\omega}$, $h_r \geq 0$ and the equilibrium condition is also fulfilled.

Krugman's model framework does not change the three main short run equilibrium equations (2.12), (2.14), and (2.16) from above. However, empirical analyses also have to take the long run equilibrium into account. Existing studies concentrate on the dispersed equilibrium outcome. Therefore, respective authors introduce an additional assumption which is given by

$$\frac{w_r}{q_r^\varepsilon} = \frac{w_s}{q_s^\varepsilon}. \quad (2.18)$$

However, this assumption is restrictive in case of empirical application, especially in the case of Europe where labor mobility is rather low and real wage equalization is unrealistic.

Extensions and Modifications. Krugman's insights were modified and extended in literature for three reasons. First, dispersion can also be incorporated by means other than a traditional sector. Second, agglomeration processes can also emerge via supply-based linkages. Third, authors wanted to overcome the unrealistic extreme agglomeration outcome of the original model.

One popular variant of the first strand is developed by Helpman (1998), who substitutes the traditional sector of the Krugman-model by a housing sector. The good provided by the housing sector can not be traded amongst the regions. Consequently, the housing price is not identical among regions, but cheaper in the periphery where housing demand is lower. This fact inverts a conclusion of the Krugman-model: Decreasing transport costs now facilitate the access to industrial goods in the periphery and thus favor dispersion. In this model, there are another three relations which need to be fulfilled in order to achieve the equilibrium (see Helpman (1998) for details.) First, regional income needs to equal the income derived from labor and housing. Second, the market value of the housing services supplied equals the share of the income spent on housing services:

$$P_r H_r = (1 - \varepsilon) Y_r, \quad (2.19)$$

where P_r is the housing price in region r and H_r is the fixed stock of housing in region r . Third, the market for traded goods needs to be in equilibrium. Real wages are now given by $\frac{w_r}{P_r^{1-\varepsilon} q_r^\varepsilon}$. Therefore, assumption (2.18) changes to

$$\frac{w_r}{P_r^{1-\varepsilon} q_r^\varepsilon} = \frac{w_s}{P_s^{1-\varepsilon} q_s^\varepsilon}. \quad (2.20)$$

The second strand follows Venables (1996). He shows that the role of mobile workers for the endogenous determination of the size of regional markets could be replaced by vertical linkages between up- and downstream industries. He assumes three sectors: one traditional and two modern sectors.⁴ Both modern industries face economies of scale. This allows for input-output linkages between firms while ruling out inter-regional migration. When firms are attracted by another region and are relocating, they carry on their demand for intermediate goods with them and thus create agglomeration processes. In this model, the wage equation should be called more accurately an equation for the price of the composite immobile factor of production (Redding & Venables (2004: 58)). Puga (1998) generalizes the wage equation for the vertical linkages model. Due to the inclusion of intermediate goods,

⁴For the sake of simplicity, Krugman & Venables (1995) merge the two modern sectors to one.

equation (2.16) now consists of the terms stated above plus the value of demand for intermediate inputs and price q_s :

$$w_s = \phi q_s^{-\gamma/(1-\gamma)} \left[\sum_r \left(\frac{\tau_{sr}}{q_r} \right)^{1-\sigma} D_r \right]^{1/(\sigma(1-\gamma))}, \quad (2.21)$$

where ϕ collects constant terms, γ is the share of intermediates in the production process, and $D_r = \varepsilon Y_r + \gamma h_r p_r \bar{x}$.⁵ However, this model shares the features of the Krugman-model since it exhibits the same equilibrium properties.

The third strand of modifications was introduced by Puga (1998). He combines the approaches of Krugman (1991) and Krugman & Venables (1995). Assuming interregional immobility of workers, decreasing levels of transport costs do not automatically lead to an agglomerated outcome as in the Krugman case. Agglomeration tendencies are weakened by the fact that a growing home market can only be realized if workers are attracted from other sectors of a region to the industrial sector. An intersectoral shift causes rising wages which leads to increasing production costs in the agglomeration. The difference in production costs grows bigger with lower transport costs. This is the reason why in case of regional immobility of workers, dispersion is the equilibrium outcome for very low levels of transport costs. This means catastrophic agglomeration is replaced by gradual agglomeration and non-extreme partial agglomeration outcomes become possible.

2.2 Survey of Empircis

In this section, we present two different approaches towards the questions raised in the beginning. To which extent does regional accessibility explain regional income differences? And can we rely on a fully parameterized model?⁶ First, we will discuss the method proposed by Hanson (2005). He takes the wage equation as starting point for his analysis. Second, we present the two stage approach of Redding & Venables (2004). In the first stage, they estimate the gravity trade relation in order to recover accessibility measures. Their second step is using these generated indicators in an estimation of the wage equation.

⁵See Fujita *et al.* (2001: 241ff), for details. Note that in the Krugman- and Helpman-models $\gamma = 0$.

⁶There are far more studies which isolate estimates for σ . These studies follow a different strategy relying entirely on the gravity equation. The most prominent ones are Feenstra (1994), Hummels (1999), Haveman & Hummels (1999), Head & Ries (2001), Limao & Venables (2001), Baier & Bergstrand (2001) and Lai & Trefler (2002). See Anderson & van Wincoop (2004) for an extensive survey.

2.2.1 The Hanson Approach

The review of empirical studies begins with the so-called Hanson-approach. Hanson (2005) tries to examine the Helpman-model via the wage equation. In order to obtain a testable relation, he substitutes equilibrium conditions (2.20) and (2.19) into the wage equation (2.16). Hence, the author assumes perfect mobility of labor – which yields equilibrated real wages across regions and levels a steady state status of the analyzed regions. Furthermore, he assumes equality of revenues and expenditures in the housing sector. This leads to a wage equation which does not directly rely on regional price indices. This is necessary as these are not available in statistics data bases. For region s , q_s is represented by an expression referring to the income level in region r (Y_r), the housing stock (H_r), and wage levels (w_r) in all other regions r :

$$q_r^{\sigma-1} = cY_r^{\frac{(1+\sigma(\varepsilon-1))}{\varepsilon}} H_r^{\frac{(\sigma-1)(1-\varepsilon)}{\varepsilon}} w_r^{\frac{(\sigma-1)}{\varepsilon}}, \quad (2.22)$$

where c is a constant (see Hanson (2005) for details). Using this expression in equation (2.16), taking logs and adding an additive error gives the empirical specification of the wage equation:

$$\log(w_s) = \theta_0 + \theta_1 \left[\log \left(\sum_r Y_r^{\theta_3} w_r^{\theta_4} H_r^{\theta_5} \exp(\theta_2 g_{sr}) \right) \right] + \nu_s, \quad (2.23)$$

where w_s is the average annual wage, g_{sr} is the distance between regions s and r , and ν_s the vector of regression errors. Transport costs are specified according to the exponential function $\tau_{sr} = \exp(\zeta g_{sr})$, where ζ is the distance parameter. According to equation (2.22), the regression coefficients θ are entirely determined by the model parameters. That is why it is possible to achieve a fully parameterized model when estimating equation (2.23). Note, that the author does not use the other two equilibrium conditions mentioned above. Hence, this strategy cannot be considered a full-information approach of NEG theory.

Results for the U.S. Hanson (2005) proposes to estimate two specifications of equation (2.23). The first is the baseline specification, in which coefficients are restricted to $\theta_3 = 1$, $\theta_4 = 0$ and $\theta_5 = 0$. The resulting formula reminds of the traditional market potential function by Harris (1954). The second and unrestricted specification from equation (2.23) is referred to as the augmented market potential formula. He estimates both specifications in first differences in order to eliminate region specific effects. However, one econometric problem arises from equation (2.23). Regional wages appear both on the left and the right hand side. Additionally, wages

are implicitly included in personal income. This leads to endogeneity problems when estimating. To study the impact of endogeneity, Hanson uses the econometric techniques of Nonlinear Least-Squares (NLS) and the Generalized Method of Moments (GMM). In the latter, he can correct for endogeneity. He instruments the “regressor function”, hence the terms in squared brackets in equation (2.23), with county population growth, lagged by 10 years. This strategy accounts for hypothetical correlation between the error term and changes in the “regressor function”.

The author can rely on regional manufacturing wage data as well as on regional housing stock statistics for 3075 U.S. counties. He approximates Y_r by county GDP levels. However, right hand side variables in brackets are not measured on the county level due to computational problems. Instead, Hanson aggregates surrounding counties within concentric distance bands.⁷ This procedure should also weaken endogeneity problems when estimating.

Hanson’s results concerning the differences of time spans 1970-1980 and 1980-1990 are as follows. In both specifications tested, he verifies that geographical concentration of economic activity is due to product-market linkages between regions which result from economies of scale and transport costs. Stated differently, the author finds that regional variation in wages is associated with proximity to large markets. However, these strong demand linkages are limited in geographic scope since the coefficient for distance θ_2 is relatively high. This means that the distance decay of transport costs is rather steep. Nevertheless, Hanson (2005) rejects the simple model in favor of the augmented market-potential model. The values for the elasticity of substitution are in plausible ranges (i.e. $1.7 \leq \sigma \leq 2.5$). Estimation results are not influenced by the inclusion of wage controls such as human capital externalities or exogenous amenities, which could be alternative explanatory variables for regional wage differences. But Hanson (2005) admits that his set of control variables is not exhaustive. Furthermore, the author entirely ignores spatial econometric problems in his setting, not even testing for their existence, although it is most likely that the data exhibit spatial autocorrelation on county level.

Results for Europe. The Hanson approach has already been applied to the European data as well. We concentrate on three studies. The first uses the Krugman-model, the second relies on the Helpman-model, and the third is based on the Puga-model.

⁷The bandwidth increases with increasing distance.

- In a first study, Niebuhr (2004) applies the wage equation of the Krugman-model and compares it to a simple market potential function. Because there is a lack of data on price indices, she assumes $q_r = q_s$. Using this in equilibrium condition (2.18) gives the nominal wage equation in logs:

$$\log(w_s) = \theta_0 + \theta_1 \log \left(\sum_r Y_r w_r^{\theta_4} \exp(\theta_2 g_{sr}) \right) + \nu_s, \quad (2.24)$$

where transport costs are modelled according to an exponential distance function. The market potential function does not take into account regional wages on the right hand side, thus $\theta_3 = 0$.

She estimates cross-sections for 1985, 1990, 1995 and 2000 in levels for a data set of 158 Western European NUTS2⁸ regions. They are estimated by NLS and Spatial Maximum-Likelihood (SML). Niebuhr (2004) can directly rely on wage data. Moreover, she uses proxies for inner regional distances of the form $g_{rr} = 3/4\sqrt{area_r/\pi}$, where *area* is a region's size in km^2 (see Head & Mayer (2000) for a discussion of the alternatives for modelling internal distances).

The model's predictions are verified over all specifications, and estimates for the elasticity of substitution range between $5.65 \leq \sigma \leq 12.01$. Although the values for σ are higher than in other studies, they are still consistent with the theory. They rise with the inclusion of wage controls, namely industry structure, natural amenities and population density.

However, applying difference-in-difference techniques yields implied values for σ that are partially negative, which is counterintuitive and not consistent with the theory. Estimations including measures for border effects⁹ are also disappointing since these effects have a statistically insignificant impact. Unfortunately, estimations which solve endogeneity issues in the style of Hanson (2005) turn out to be unstable and are not reported in the original study.

- Second, Mion (2004) elaborates the Helpman-model for 103 Italian provinces (NUTS2 level). He uses dynamic panel techniques à la Arrelano & Bond (1991) in order to control for the fact that most variables in NEG are endogenous. To

⁸See Eurostat (1999) for details on the classification.

⁹The evaluation is established by adding a matrix of border dummies to the distance matrix g_{sr} , thus additionally estimating one coefficient per border dummy. This might be important, as empirical studies on border effects suggests that trade decreases considerably when international borders are crossed (e.g. Bröcker (1984), McCallum (1995), Helliwell (1996), Wei (1996), Nitsch (2000), Head & Mayer (2000), Parsley & Wei (2001), Feenstra (2002) and Anderson & van Wincoop (2003)).

solve endogeneity problems, Mion (2004) instruments endogenous variables by their past levels, lagged by two years. Mion covers the time period 1991-1998. The author approximates wage data by subtracting housing spending from regional GDP and dividing the result by the active population.

Adding a time dimension t , the author initially estimates the original nonlinear setting from equation (2.23). Mion (2004) incorporates transport costs using a power function approach, hence $\tau_{sr} = \zeta g_{sr}^\kappa$, where parameters $\zeta > 0$ and $\kappa > 0$. For $0 < \kappa < 1$ the transport cost function is concave, for $1 < \kappa$ it is convex.¹⁰ The author assumes a distance decay of $\kappa(1 - \sigma) = -1$ according to estimates in Disdier & Head (2008). Results confirm the outcome in Hanson (2005) for the Italian data. It turns out that the values for σ depend on whether he instruments or not, but they range between plausible bounds of 1.9 and 6.7.

Subsequently, Mion (2004) turns to his own empirical strategy. He linearizes the problem¹¹ since nonlinear estimations with endogeneity and instrumentation issues often turn out to be unstable as this mix causes criterion functions to have many local minima. The model from (2.23) then reads as

$$\log(w_{s,t}) = \theta_0 + \sum_r \left[\left(\tilde{\theta}_3 \bar{Y}_{r,t} + \tilde{\theta}_5 \bar{H}_{r,t} + \tilde{\theta}_4 \bar{w}_{r,t} \right) g_{sr}^{-1} \right] + \nu_{s,t}, \quad (2.25)$$

where $\bar{Y}_{r,t} = Y_{r,t} \log(Y_{r,t}) / \sum_r Y_{r,t}$ results from a linear approximation used to linearize the model. $\bar{H}_{r,t}$ and $\bar{w}_{r,t}$ are calculated accordingly. $\tilde{\theta}_3 = \theta_1 \theta_3 \zeta$, $\tilde{\theta}_5 = \theta_1 \theta_5 \zeta$ and $\tilde{\theta}_4 = \theta_1 \theta_4 \zeta$ are equal to θ_3 , θ_5 and θ_4 , up to a common multiplier.

Adding time dummies dum and a time-lag of $\log(w_{s,t})$ allows to control for time dynamics. Stacking notation over s and r , the final regression equation becomes

$$\log(\mathbf{w}_t) = \mathbf{i} dum_t + \log(\mathbf{w}_{t-1}) \theta_0 + \mathbf{W} \bar{\mathbf{Y}}_t \tilde{\theta}_3 + \mathbf{W} \bar{\mathbf{H}}_t \tilde{\theta}_5 + \mathbf{W} \bar{\mathbf{w}}_t \tilde{\theta}_4 + \nu_t, \quad (2.26)$$

where \mathbf{i} is a vector of ones and \mathbf{W} is a spatial weights matrix with generic elements $W_{sr} = 1/g_{sr}$. Diagonal elements of \mathbf{W} are set equal to zero.

The linearized model estimates confirm the model's predictions. Hence, regional relative prices explain the spatial wage structure in Italy. Implied values for σ are between 3.2 and 3.4. Spatial tests show no evidence for spatial misspecification. However, no wage controls are included in this study.

¹⁰For theoretical arguments in favor of this transport cost specification, see McCann (2005).

¹¹This strategy was pioneered by Combes & Lafourcade (2001).

- The third study (Brakman *et al.* (2006)) is based on a linearized version of the model by Puga (1998). The authors also rely on the transport cost specification of $\tau_{sr} = \zeta g_{sr}^\kappa$. Internal distances are set to $g_{rr} = 2/3\sqrt{area_r/\pi}$. Equation (2.21) is thus estimated according to

$$\log(w_s) = \theta_0 - \theta_6 \log(q_s) + \theta_7 \log \left[\sum_r D_r (\theta_8 g_{sr})^{\theta_9} q_r^{\theta_{10}} \right] + \nu_s. \quad (2.27)$$

According to equation (2.21), the regression coefficients θ are again entirely determined by the model parameters: $\theta_6 = -\gamma/(1-\gamma)$, $\theta_7 = 1/(\sigma(1-\gamma))$, $\theta_8 = \zeta$, $\theta_9 = \kappa(1-\sigma)$, and $\theta_{10} = \sigma - 1$. Since it is likely to find multicollinearity between q_s and the “regressor function”, hence the term in squared brackets,¹² the authors approximate the former by a constant when estimating.

As data on price indices are not available for the EU, the authors also need to reformulate q_r contained in the “regressor function” before estimating.¹³ They approximate the price index in region r as an average of the wage in region r and the wages in the nearest center region. The latter are defined as the 35 regions which exhibit the highest market potential.¹⁴ Adjusting for the distance between regions r and the center regions, Brakman *et al.* (2006) define

$$q_r = [\varsigma_r w_r^{1-\sigma} + (1 - \varsigma_r)(\bar{w}_r g_{r,center})^{1-\sigma}]^{1/(1-\sigma)}, \quad (2.28)$$

where ς_r is region r 's share of employment in manufacturing and \bar{w} is the average wage outside district r . This reformulation leads to strong endogeneity problems which have to be accounted for when estimating.

Brakman *et al.* (2006) rely on a sample of 174 Western European NUTS2 regions. Wages on the right hand side of equation (2.27) are measured at a more aggregated regional level, hence NUTS1. D_r is approximated by regional Gross Value Added (GVA) and is measured on NUTS1 level, too. Estimations are carried out for the time period of 1992-2000 via pooled weighted least squares (WLS). The authors include dummies for border regions, river regions, and coastal regions, as well as metric variables for regional solar irradiation as wage controls. They run level and first difference regressions as well as instrumental variable (IV) estimations. However, the authors do not apply spatial econometric models in order to account for spatial misspecification.

¹²See Redding & Venables (2004) and Knaap (2004) for an extensive discussion.

¹³This strategy is based on Brakman *et al.* (2004) who estimate equation (2.23) from the Helpman (1998) model for 441 German districts (NUTS3 regions).

¹⁴See Brakman *et al.* (2006) for details.

Instrumentation is carried out using a two step least squares (2SLS) procedure. In the first step, income is regressed upon its initial 1991-level, a country-specific time trend, country dummies, and a border dummy. Then, they project income using the estimated coefficients and use the projection in the estimation of equation (2.27). In all specifications, the authors verify the model predictions and estimate values of $\sigma = 3.85$ and $\kappa = 0.35$.

2.2.2 The Redding-Venables Approach

The second empirical approach is the one issued by Redding & Venables (2004). The authors also aim to test the causal relationship between income differences and relative prices by means of a wage equation. In contrast to the studies presented earlier, their strategy does not permit a full parameterization of the underlying NEG model. However, the advantage here is the possibility to substitute price indices with estimated measures for the market and supply accesses, MA and SA , which are derived from bilateral trade relations.

In the context described above, we obtain from equation (2.14)

$$X_{sr} = A_s \tau_{sr}^{1-\sigma} B_r, \quad (2.29)$$

where $A_s = h_s p_s^{1-\sigma}$ and $B_r = q_r^{\sigma-1} D_r$.

By summing equation (2.29) over r we obtain

$$p_s^\sigma \frac{X_{s\bullet}}{p_s h_s} = \sum_r \tau_{sr}^{1-\sigma} B_r, \quad (2.30)$$

where $X_{s\bullet} := \sum_r X_{sr}$. The left hand side of equation (2.30) is defined as market access MA_s . Following the Dixit-Stiglitz-model, the number of variants h_s is proportional to the real output $X_{s\bullet}/p_s$. The fraction on the left hand side of equation (2.30) therefore is a constant and hence MA_s equals p_s^σ up to a constant. The higher p_s , the higher is the market access for region s .

By summing equation (2.29) over s we get

$$X_{\bullet r} = D_r q_r^{\sigma-1} \sum_s \tau_{sr}^{1-\sigma} A_s, \quad (2.31)$$

where $X_{\bullet r} := \sum_s X_{sr}$. Because in equilibrium $X_{\bullet r} = D_r$, we obtain

$$q_r^{1-\sigma} = \sum_s \tau_{sr}^{1-\sigma} A_s. \quad (2.32)$$

The left hand side of equation (2.32) is defined as supply access SA_r . The smaller q_r , the higher is SA_r .

MA_s and SA_r can be specified empirically by estimating equation (2.29) based on trade data and then calculating MA_s and SA_s for every single region. This is done via the following formulas:

$$MA_s = \sum_r \tau_{sr}^{1-\sigma} B_r \quad (2.33)$$

and

$$SA_s = \sum_r \tau_{rs}^{1-\sigma} A_r. \quad (2.34)$$

Redding & Venables (2004) specify τ_{sr} as

$$\tau_{sr}^{1-\sigma} = g_{sr}^{\theta_{11}} bord_{sr}^{\theta_{12}}, \quad (2.35)$$

where

$$bord_{sr} = \begin{cases} 0 & \text{if } r \text{ and } s \text{ are neighboring regions} \\ 1 & \text{else} \end{cases} \quad (2.36)$$

and θ_{11} as well as θ_{12} are regression coefficients in the logarithmic ansatz of equation (2.29).

In a second step, the access measures are related to the wage equation in Krugman & Venables (1995), which comprises supply and demand prices (see section 2.1.2 and Redding & Venables (2004) for details). The wage equation is now written in logs with regression errors ν_s as

$$\log(w_s) = \theta_0 + \theta_{13} \log(\widehat{SA}_s) + \theta_{14} \log(\widehat{MA}_s) + \nu_s. \quad (2.37)$$

where \widehat{MA}_s and \widehat{SA}_s are market and supply accesses calculated in step one. Unfortunately, $\hat{\theta}_{13}$ and $\hat{\theta}_{14}$ are not identified if \widehat{MA}_s and \widehat{SA}_s are both used in the estimation of equation (2.37). Therefore, Redding & Venables (2004) propose to rely on only one of the two market access indicators when estimating.

International Results. Redding & Venables (2004) study a data set of 101 countries. Hence, region indices s and r in equations (2.29)-(2.37) are replaced by country indices l and k . They estimate equations (2.29) and (2.37) for cross-sections of the year 1994. In the first step estimation, intranational trade is calculated as domestic production minus exports (see Wei (1996) for details). The authors try to

solve the zero observation problem for bilateral trade flows by adding 1 to all trade flows before taking logs.¹⁵ Redding & Venables (2004) control for the left-censored trade matrix applying a Tobit estimator. Concerning transport costs they impose a power function specification of the form $g_{lk}^{\theta_{11}}$.¹⁶ Internal distances g_{kk} are measured as $g_{kk} = 2/3\sqrt{(area_k/\pi)}$, where $area_k$ is the area of country k in km^2 . In the second step estimation, the authors proxy the left hand side variable with incomes per capita and apply OLS. They need to use bootstrapping methods in order to obtain valid standard errors since constructed regressors are used.

Redding & Venables (2004) find that the model's predictions are confirmed. A good market access significantly affects the income level in a positive way and gives rise to international income differences. Robustness checks include wage controls such as diverse measures for natural amenities and for political issues. Furthermore, they instrument the accessibility measure with distance to the economic centers, such as the United States, Belgium, and Japan. These modifications do not change the implications of the results.

Results for the U.S. On the regional level, the method by Redding & Venables (2004) is applied to a sample of 2061 U.S. counties by Knaap (2006). He estimates the gravity relation (2.29) using data from 1997. Internal distances are calculated as $g_{rr} = \sqrt{area_r/\pi}$. For the estimation of (2.37), wage data are taken from the year 1999. The author identifies significant relations between regional wages and market access. However, this relation turns out to be insignificant after adjusting for estimation biases by including first nature wage controls. Additionally, Knaap (2006) tests for spatial autocorrelation in the data but finds no significant distortions. He also applies the Redding-Venables instrumentation strategy which confirms the results. However, the findings suggest a limited role of economic geography in the explanation of wage differences in the U.S.

Results for Europe. Other than for the U.S., researchers can not rely on regional trade data when they are interested in examining the European regions. To our best knowledge, only two attempts have been undertaken to apply the Redding-Venables approach to European trade data.

- The first is carried out by Head & Mayer (2006), who estimate a wage equation

¹⁵This correction is controversial. For instance, Bröcker & Rohweder (1990) propose to estimate gravity trade relations nonlinearly using Quasi-Maximum-Likelihood techniques to overcome the zero observation problem.

¹⁶This specification is also applied by all subsequent authors who use the described approach in their analysis.

for 57 Western European NUTS1 regions for the time period 1985-2001. In the first step estimation, the authors implement a second border effect, namely an interaction term consisting of a border and a language dummy. Internal distances are measured as in Redding & Venables (2004). Although Head and Mayer are interested in the regional context, they estimate the trade relation (2.29) on the national level. Assuming that regional trade flows are governed by similar underlying forces as national trade flows, they take the national estimates for market and supply accesses and disaggregate them to the regional level, using regional incomes as weights. Obviously, this approach is problematic. Assuming that the first step estimation à la Redding & Venables (2004) were valid, the authors destroy the gathered information by almost arbitrarily allocating the estimated accessibility measures among regions. It would be more appropriate to eliminate the first step estimation and then simply use atheoretic accessibility measures – such as the one by Harris (1954) – in the (second step) estimation.

In their second step estimation of equation (2.37), Head & Mayer (2006) can rely on wage data as they opt for an aggregated data set on the NUTS1-level. The authors admit that their left hand side variable (wage per employee) lacks of comparability across countries as the reporting standards differ considerably. Consequently, they use a pooled regression with country dummies in order to capture these inconsistencies. Furthermore, Head and Mayer apply instrumental variable estimations. They use the distance to the economic center of Europe and the centrality, defined as $\log \sum_s g_{sr}^{-1}$, as instruments. The latter is chosen because it is independent of the market sizes of regions. The only wage control used is average years of education. In all settings, Head and Mayer validate a significant impact of market proximity on regional wages. Finally, the implied values for the elasticity of substitution range between $8.9 < \sigma < 12.6$.

- The second attempt is by Breinlich (2006) who analyzes 193 Western European NUTS2 regions for the time period 1992-1997.¹⁷ In his first step estimation of equation (2.29), Breinlich also includes a country's trade relations with the rest of the world by incorporating a second term for border effects. The first effect is catching the trade impediments between trading partners in two different EU-countries. The second effect considers the effect of goods being shipped across the EU-border. Furthermore, he adjusts the distances between trading

¹⁷His approach is comparable to the one by Head & Mayer (2006).

nations since the capital-to-capital distances between neighboring countries overestimate the true distance relevant for trade flows (Head & Mayer (2002a)). Adjusted distances are calculated as the gross value added (GVA)-weighted distances between the main cities of the NUTS2 regions of the two trading partners k and l :

$$g_{lk} = \sum_{r \in \mathcal{I}_k} \sum_{s \in \mathcal{I}_l} \iota_r \iota_s g_{sr}, \quad (2.38)$$

where \mathcal{I}_k is the region set of country k and ι_r is a the share of region r in country k 's GVA. For the sake of simplicity, we use the same symbols for regional and national distances. For trade between countries from the EU-15¹⁸ and the rest of the world, this boils down to the sum of GVA-weighted distances between regions and non-EU-15 countries:

$$g_{lk} = \sum_{r \in k} \iota_r g_{sr}. \quad (2.39)$$

Internal distances are calculated as $g_{rr} = 2/3 \sqrt{area_r/\pi}$.

After estimating the gravity equation, he disaggregates the results obtained for the international trade relations to the regional level, using regional GVA shares as weights. Breinlich argues that his way of calculating international distances takes account of the critique raised against the arbitrary approach of Head & Mayer (2006).

Breinlich's (2006) findings on the wage equation (2.37) are based on a pooled OLS approach. Additionally, he carries out fixed effects panel data estimations. He also instruments the accessibility measures with distance to Luxembourg and the size of the region's home country. The author approximates regional wages by GVA in 1995 prices per active population. He also includes wage controls such as education levels, regional capital stocks, and population density. In all settings, Breinlich (2006) finds an important role of market access to explain regional income differences. Assuming a labor share in regional GVA of 0.6, Breinlich (2006) is able to report implied σ -values of 6.3, comparing it to a corresponding value of $\sigma = 3.3$ in Redding & Venables (2004).

Finally, we resume outcomes for the elasticity of substitution from the presented empirical studies. This provides a basis for comparing the results of our own approach with the outcomes in related literature. The summary in table 2.1 shows that for

¹⁸EU-15 comprises all EU-member countries prior to EU's Eastern Enlargement in 2004.

Table 2.1: Key Parameter Estimates in Literature

Author	Approach	Study area	Elasticity of substitution σ
Hanson (2005)	Hanson	USA	4.9–7.6
Mion (2004)	Hanson	Italy	1.9–6.7
Niebuhr (2004)	Hanson	Western Europe	5.6–12
Brakman <i>et al.</i> (2006)	Hanson	Western Europe	3.85
Redding & Venables (2004)	R&V	International	3.3
Knaap (2006)	R&V	USA	1.9–2.5
Head & Mayer (2006)	R&V	Western Europe	8.9–12.6
Breinlich (2006)	R&V	Western Europe	6.7

Europe we can conclude that value ranges are roughly $2 < \sigma < 12$. This span is quite large and provokes different implications for policy decisions. The reasons for this are that the studies described rely on different model variants, on different transport cost implementations, on different European sub-spaces, and use different econometric estimation techniques which alas all influence the estimates.

2.2.3 Critique

Following the two presented strategies by Hanson (2005) and Redding & Venables (2004), studies identify the distance to the main markets as the most important source for spatial income differences. However, these approaches use different paths in order to derive this conclusion. Redding & Venables (2004) estimate the international correlation between per capita incomes and the proximity to a nation’s import demand. Hanson (2005) studies the relation between regional wage growth and the change of the regional market potential function (i.e. demand linkages). However, both approaches have their limitations: On the one hand, Hanson neither explicitly relies on regional relative prices for manufacturing goods nor on the trade relations between regions. On the other hand, Redding & Venables cannot obtain a fully parameterized model. In order to come to a better approach, both frameworks of analysis should be merged to one.

The major shortcomings of the presented empirical literature are the following.

- Knaap (2004) argues that estimating the first and second step equations in the Redding & Venables (2004) approach as well as only estimating the wage equation by the Hanson (2005) approach means that one holds every feature of the model – else than the specific one tested – constant. That is why the studies

presented do not account for the numerous linkages in the model and they might, therefore, obtain biased estimates. Knaap (2004) proposes to estimate the entire equilibrium solution of the model instead of looking at the single equations. This can be done by means of *Computable General Equilibrium* (CGE) methods which calibrate the respective model's equilibrium solution to real world data. These outcomes can be used for estimating the wage equation.

- The authors using the Hanson approach do not properly solve the problem of unobservable regional price indices. They impose restrictive assumptions of equalized real wages or equalized price indices to get rid of them in their regression equations (Hanson (2005), Niebuhr (2004)). Alternatively, they approximate expressions for price indices in order to escape the problem of data shortages (Brakman *et al.* (2006)). Both strategies lead to endogeneity problems because regional wages appear on the left and on the right hand side of the wage equation. To our best knowledge, there are no attempts to calibrate price indices via usage of CGE methods in order to apply them directly in the estimation of the wage equation.
- Regional trade relations can neither be retrieved from statistics databases for Europe. This problem is not satisfyingly solved in the reviewed empirical studies. As described, several authors tried to make use of the Redding-Venables approach for Europe, allocating national trade data to the regional level using arbitrary weights (e.g. regional GDP shares). Calibrating a CGE trade model would allow to obtain regional trade relations and to avoid the use of such arbitrary techniques.
- The authors use different functional forms for transport costs, hence the exponential or the power function. Moreover, some authors using the Hanson approach do only rely on distance measures but do not include trade impediments such as border effects into their analysis. Since the importance of border effect for international trade is widely proven, it would be preferable to apply a more sophisticated measure for border effects. This can also be done by calibrating a CGE model that explicitly accounts for bilateral trade impediments.
- A further concern with respect to Europe is, that the incomplete study area yields biased results. Studies which comprise regions of the entire EU-27 are non-existent, but only such a frame would take into account all linkages on the continent. In literature, we can only find studies that focus on Western

Europe or on single nations. Especially, no evidence for Eastern Europe based on structurally derived NEG equations has yet been retrieved.¹⁹

- The econometric methods applied in estimation are debatable. One should take spatial correlation, endogeneity, control variables, and region specific effects into account. This has not always been accomplished in the framework of a single econometric approach.

2.3 Chapter Summary

In this chapter, we have shown that the state of the art literature provides evidence that regional market accessibility seems to have a significant impact on the regional income distribution. We presented the two empirical strategies by Hanson (2005) and Redding & Venables (2004) and discussed their shortcomings. Hanson (2005) already stated that it would be appealing to combine both approaches described in order to allow consumers and producers to be a source of industrial demand. At the same time, the combined approaches would make it possible to recover all structural parameters of the model analyzed. However, neither the necessary interregional trade nor the regional price data are available, at least for Europe.

To our best knowledge, no author tried to overcome these restrictions and to apply the Hanson approach to a model which allows to calculate market proximity, e.g. regional relative prices, directly. This is our goal in the present monograph. We use a CGE model and combine it with an econometric estimation approach.²⁰ The gravity equation is not explicitly estimated but used to calibrate the CGE model and to obtain regional trade measures, trade impediments and regional equilibrium prices for each time period considered. We thus do not need to apply restrictive assumptions of identical price indices or equalized real wages. We allow equilibrium real wages to vary among regions because we do not assume perfect mobility of

¹⁹Nevertheless, there are few studies which estimate relations resembling to NEG wage equations. However, model predictions cannot be verified nor can parameter values be retained from these approaches. The only purpose is to detect whether concentration or specialization patterns influence interregional income differences. Hence, existing studies are more in the tradition of Kim (1995), Ellison & Glaeser (1997), Amiti (1999), and Ellison & Glaeser (1999). See e.g. Damijan & Kostevc (2003) or Brühlhart *et al.* (2004) for studies for Eastern Europe.

²⁰However, there are several attempts in literature to combine CGE and estimation approaches. Knaap (2004) applies a CGE model to his estimation à la Redding-Venables to assure that the estimated model is in equilibrium. Therefore, he needs to calibrate regional trade flows that correspond to the equilibrium solution and reproduce observed trade flows as exactly as possible. Redding & Sturm (2005) combine CGE methods and an econometric approach for a different field of study. Unsatisfyingly, the latter needed to assume a value for σ rather than being able to estimate it. However, none of these studies calibrates regional prices directly.

workers in our model. Furthermore, our study relies on a far broader data set which comprises the entire EU-27 on a disaggregated regional level (NUTS3), accounts for the econometric problems mentioned (endogeneity, spatial correlation, region specific effects), and controls for alternative forces explaining the regional income distribution.

Chapter 3

The Analytical Framework

In this chapter we turn to the proposition of our model. We apply a fully specified general equilibrium model to real world data in order to overcome the described shortcomings of existing empirical studies. In detail, we aim at simultaneously calibrating regional trade flows, border impediments and regional demand and supply prices as well as estimating a wage equation which is nonlinearly dependent on some of the model's parameters. This appealing combination of calibration and estimation techniques allows us to rely on the entire equilibrium solution of our model and to dispense of using arbitrary techniques of disentangling regionally aggregated data.

We use the *CGEurope Model*, which was conceptually introduced by Bröcker (1995).¹ It is closely related to the models described in chapter 2. Therefore, a comparison of our model's outcome to those of the approaches reviewed earlier is possible. However, the simultaneous test of our model's key equations should augment the reliability of empirical results compared to the studies that concentrate on single model relationships.

The chapter is organized in two sections. Section 3.1 gives an introduction to the model's setup. In section 3.2 we derive our empirical strategy. Specifically, it is shown how we calibrate the model and how we obtain the empirical specification of the wage equation.

¹It was computationally implemented by Bröcker (1998b). Applications of the models comprise the quantification of regional welfare effects due to European integration patterns such as declining trade barriers (Bröcker (1998b)), infrastructure investments (Bröcker (1998a)) and the European Monetary Union (Bröcker (2003)).

3.1 The *CGEurope* Model

In this section, we present the model's assumptions and its analytical framework.² We first concentrate on the analysis of the supply and demand side as well as on modelling the transportation sector. Concluding, the equilibrium is presented.

3.1.1 Assumptions

We build our analysis on a static two-sector general equilibrium model of monopolistic competition in the tradition of Dixit & Stiglitz (1977). The two production sectors are a tradables and a non-tradables sector. In the tradables sector the firms face a market structure of monopolistic competition, whereas the producers in the non-tradables sector interact according to the rules of perfect competition. Furthermore, we assume perfect price flexibility, which means that all markets are cleared. Consequently, the economy is under full employment. The spatial dimension is introduced through the assumption that transport of goods is costly. These transport costs consist of a distance-related and an impediment-related component. As usual, we concentrate on the modelling of the monopolistic competition sector. Note that in contrast to the NEG models mentioned before, we do not rely on endogenous agglomeration processes.

3.1.2 Supply Side

Firms are maximizing profits following Dixit & Stiglitz (1977). The lower-tier of the maximization problem determines the economy's provision with tradable goods whereas the upper-tier gives the overall production. Let us turn to the lower-tier problem first. According to the theory of monopolistic competition there is a large range of varieties of a certain tradable good. The single producer's profit maximization problem for determining the minimal cost of production of the respective variety is straightforward.³ Since we have no specific information about the individual varieties, they enter symmetrically into the cost function. Hence, aggregated over all producers of tradables, we can write the minimal unit cost of a composite tradable as

$$q_r = \phi \left(\sum_{i=1}^{\bar{h}} k_{r,i}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \quad (3.1)$$

²This section is based on Bröcker (1998a).

³See chapter 2.

where q_r is the cost index for the aggregate demand of tradables in region r , $k_{r,i}$ describes the cost of a single variety i used in region r , and σ denotes the constant elasticity of substitution. Finally, \bar{h} is the overall number of varieties produced and ϕ is a constant.

The upper-tier problem determines the supply price p_r of a tradable good produced in region r . This price consists of three components: the cost for factor services w_r ,⁴ local inputs costs p_r , and intermediate tradables costs q_r . All factors used in production are entirely owned by the households. The sensitivity of p_r with respect to a variation of the input costs is described by elasticities α , β , and γ . Note that $\alpha + \beta + \gamma = 1$. The costs of the input-mix are weighted by the productivity level m_r . Formally, we therefore have as minimal unit cost the following Cobb-Douglas relation:

$$p_r = \frac{1}{m_r} w_r^\alpha p_r^\beta q_r^\gamma. \quad (3.2)$$

Solving equation (3.2) for w_r gives the equation for the price of the factor service:

$$w_r = m_r^{\frac{1}{\alpha}} p_r^{\frac{1-\beta}{\alpha}} q_r^{-\frac{\gamma}{\alpha}}. \quad (3.3)$$

The regional factor income Y_r is defined by the product of factor price w_r and the exogenously given fixed regional factor stock K_r :

$$Y_r = w_r K_r. \quad (3.4)$$

Taking the above considerations into account, we can determine the output quantities of the local goods. A profit maximizing firm sets the price for the good produced equal to the product of marginal costs and a mark-up factor of $\sigma/(\sigma - 1)$, where $\sigma > 1$. In monopolistic competition markets, firms have free market entry. Therefore, this price also equals the firms' average cost. Expressing this analytically yields:

$$\frac{I(x)}{x} p_r = \frac{\sigma}{\sigma - 1} \frac{dI(x)}{dx} p_r, \quad (3.5)$$

where x is the output of a specific tradable good and $I(x)$ the input required to produce x . Solving equation (3.5) for x gives the profit maximizing output \bar{x} of a firm as a function of σ . Note that any variation of region's tradable output is due to a variation in the number of variants. The quantity of each variant remains the same.

⁴For the moment, factor service is modelled as service of a single homogenous factor – we will relax this assumption in section 3.2.

From the profit maximizing output, we can determine the supply value of varieties stemming from region r . This supply value (S_r) equals the product of the optimal input quantities of one variety $I(\bar{x})$, multiplied by the number of varieties produced (h_r), and this product being valued at the tradables unit free on board (f.o.b.) price p_r . We have

$$S_r = h_r I(\bar{x}) p_r. \quad (3.6)$$

Alas, the number of varieties produced is proportional to the real output of tradables.

For later calibration purposes, we need to relate the supply value to income and to isolate the supply value which is available for trade with other regions. Therefore, we define the two crucial components of regional supply, namely the gross regional output Q_r and the final demand of goods N_r . From equation (3.2) we know that a fraction β of the gross regional output Q_r itself is used up in production. Due to Cobb-Douglas-preferences with respect to local and tradable goods, Q_r can be defined as a fraction of regional GDP, namely Y_r/α . N_r equals the regional income Y_r plus a received net transfer G_r . These transfers are determined by the trade balance of each region which is described below. If $(1 - \varepsilon)$ denotes a constant share of consumption for local goods, the amount of $(1 - \varepsilon)N_r$ is used up in consumption of the local good. Analytically, we obtain for the value of supply:

$$\begin{aligned} S_r &= Q_r - \beta Q_r - (1 - \varepsilon)N_r \\ &= \frac{1 - \beta}{\alpha} Y_r - (1 - \varepsilon)N_r. \end{aligned} \quad (3.7)$$

Hence, the value of tradables supply of a region equals the value of production diminished by the local goods fraction of intermediate and final demand.

3.1.3 Demand Side

Via derivation of demand functions from the utility maximization problem described in section 2.1.1, we can state an expression for the value of demand:

$$\begin{aligned} D_r &= \varepsilon N_r + \gamma Q_r \\ &= \varepsilon N_r + \frac{\gamma}{\alpha} Y_r. \end{aligned} \quad (3.8)$$

This expression is derived in analogy to the one on the supply side. Hence, the first component of D_r , i.e. εN_r , is the value of final demand for tradable goods by the consumers, expressed as fraction of the disposable income N_r . The second component is the demand for the tradable good by firms.

3.1.4 Transportation and Trade

At this point we turn to the interaction between regions and show how goods are exchanged. First, we focus on the modelling of transportation, then we derive the gravity trade relation.

Transportation. If a good is shipped from one location to the other, transport is costly. In order to account for this, Bröcker (1998b) does not choose the iceberg form of transport costs that was introduced by Samuelson (1952). By Bröcker's assumption, a certain amount of the tradables composite is required for performing the transport service. We assume that transport costs are paid to a transport service. This service consumes composite tradables in the same way firms and households do and it is a non-profit industry. Transport costs do not only include physical transport of goods but also communication costs between seller and buyer, travel costs generated by maintenance services, and all kinds of non-tariff barriers to trade (Bröcker (1998a: 96)).

Following these assumptions, we can now specify the variety cost $k_{r,i}$ from equation (3.1):

$$k_{r,i} = \frac{I(\bar{x})}{\bar{x}} p_s \tau_{sr}, \quad (3.9)$$

where I is the input, expressed as a function of the optimal output quantity \bar{x} . Thus, a variety's production cost is described by the optimal input-output ratio $I(\bar{x})/\bar{x}$, valued at unit c.i.f. prices ($p_s \tau_{sr}$). The latter consist of the unit f.o.b. price for a good from region s (p_s) and the cost $\tau_{sr} \geq 1$ that arises when the good is shipped from region s to region r . Variety cost $k_{r,i}$ enters h_s times into the price index (3.1). Using equation (3.6), we have

$$\begin{aligned} q_r &= \phi \left(\sum_s h_s \left(\frac{I(\bar{x})}{\bar{x}} p_s \tau_{sr} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \\ &= \mu \left(\sum_s S_s p_s^{-\sigma} \tau_{sr}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \end{aligned} \quad (3.10)$$

where μ collects all constants in a new parameter.

Now, a rigorous specification of the transport costs is necessary. This means we not only have to formulate a proper specification of transport costs and their single components but we also have to find a precise and at the same time flexible functional form. With respect to the first point, Bröcker (1998b) assumes that transport costs consist of two components, one related to regional distance g_{sr} and another related to trade impediments between countries k and l , $z_{lk} \geq 1$.⁵ Here, the regional distance cost factor has a certain functional form $f(\cdot)$, with $f(0) = 1$:

$$\tau_{sr} = f(g_{sr})z_{lk}. \quad (3.11)$$

For $f(\cdot)$ we could chose

$$\tilde{f}(g_{sr}) = 1 + \zeta(g_{sr})^\kappa,$$

where ζ and κ are parameters with properties $\zeta > 0$ and $0 < \kappa < 1$. Including κ means that we do not impose the restriction of choosing a power function or exponential function ex-ante but let the data decide on the shape of transport costs. However, by assuming $0 < \kappa < 1$ we restrict ourselves to concave curvatures. As we require the functional form to remain mathematically tractable, we have to think about the case $\tilde{f}(g_{sr}) \approx 1$, where the transport cost function turns out to be $\tau_{sr} = \tilde{f}(g_{sr})z_{lk} \approx \exp[\zeta(g_{sr})^\kappa]z_{lk}$ (Bröcker (1998a: 11)). In order to obtain an exact specification we therefore choose:

$$f(g_{sr}) = \exp[\zeta(g_{sr})^\kappa]. \quad (3.12)$$

Nevertheless, we have to be cautious in later empirical implementations because $f(g_{sr})$ is not globally concave as it should be by assumption. For large numbers of ζ – given a certain value of $0 < \kappa < 1$ – we would have concerns as the function is not concave anymore and thus does not exhibit plausible features.

Trade. Next, we develop the trade relationship between regions. In a first step, $d_{r,i}$ is defined as the demand for variety i per unit of composite tradables. It is derived via Shephard's lemma⁶ and can be written as

$$d_{r,i} = \frac{\partial q_r}{\partial k_{r,i}} = \phi^{1-\sigma} \left(\frac{q_r}{k_{r,i}} \right)^\sigma, \quad (3.13)$$

⁵We have $z_{lk} = 1$ for $l = k$ and $z_{lk} > 1$ if $k \neq l$.

⁶See Varian (2006).

where ϕ is a constant. Then, we define V_r as the total demand in region r for tradables in real terms, including the demand for the transport service.

The value of trade X_{sr} flowing from sending region s to receiving region r can now be written as

$$X_{sr} = p_s h_s d_{r,i} V_r, \quad (3.14)$$

where the right hand side consists of two demand components, multiplied by the number of varieties, and valued at production prices. Plugging (3.9) and (3.10) into (3.13) and the result of this into (3.14) gives the relationship for the value of trade flows. It includes supply and demand prices, transport costs, and levels of regional expenditures for tradable goods. This relationship can be simplified defining a term M_r that collects variables with index r or without any index. Of course, this term remains unquantifiable for the moment, because we have not specified every single variable entering it. We will return to this problem in the next section. Keeping in mind that $S_s = p_s h_s I(\bar{x})$, we can summarize the value of trade flows as

$$X_{sr} = S_s (p_s \tau_{sr})^{-\sigma} M_r. \quad (3.15)$$

In order to fully specify the trade relation we turn to the description of the model's equilibrium.

3.1.5 Equilibrium

In equilibrium, row sums of the trade matrix have to equal the value of supply, and column sums have to equal the value of demand:

$$\sum_s X_{sr} = S_r, \quad (3.16)$$

$$\sum_r X_{sr} = D_s. \quad (3.17)$$

The latter relationship helps us finding the equilibrium trade flows in the model. Recall equation (3.15) and keep in mind that according to equation (3.17) the total value of flows to region r from all other regions s at Mill prices equals the value of demand for tradables in r at prices including transport costs but excluding demand for the transport service. This gives us the following expression for the unknown M_r from equation (3.15):

$$M_r = \frac{D_r}{\sum_s S_s (p_s \tau_{sr})^{-\sigma}}.$$

Using this expression in (3.15) gives us the regional trade relation

$$X_{sr} = \frac{S_s(p_s\tau_{sr})^{-\sigma}}{\sum_s S_s(p_s\tau_{sr})^{-\sigma}} D_r. \quad (3.18)$$

If trade is balanced, there are no transfers paid. In this case we have $N_r = Y_r$. Therefore, $D_r = S_r = (1 + \varepsilon + \gamma/\alpha)Y_r$. However, this is a purely theoretical case. In reality, trade is usually unbalanced, e.g. because of flows of capital and factor incomes or because of trade in services. Since we want our equilibrium to reproduce real trade flows we have to assume the following simple transfer scheme. Whenever a region exhibits a trade surplus it obtains a negative transfer, whereas regions with a trade deficit receive positive transfers. If transfers are paid, a region's disposable income is equal to $N_r = Y_r + G_r$.

The model's equilibrium solution depends on the parameters σ , ζ and κ .⁷ We will pay particular attention to these parameters in the next section.

3.2 The Empirical Specification

Now, we derive the empirical specification of the gravity and the factor price equations (3.18) and (3.3).⁸ First, we describe how we calibrate our model. The calibration is based on data on international trade (T_{lk}), interregional distances (g_{sr}), and regional income levels (Y_r). In equilibrium, elasticities α , β , γ , and ε correspond to the expenditures shares. Therefore, we can calibrate them with real world data. We will discriminate the shares between countries. However, for the sake of notational simplicity we do not use country indices here. Calibration gives us regional supply (p_r) and demand prices (q_r). Prices p_r and q_r are the major components of the factor prices (w_r), as stated in equation (3.3). Second, we elaborate the empirical specification of the factor price equation. It is used for two purposes. On the one hand, we are able to evaluate the role of market accessibility when explaining regional income differences. On the other hand, we can quantify parameters σ , ζ , and κ .

3.2.1 Calibration of the Model

The calibration procedure is based on the gravity equation (3.18) and a couple of constraints. For the sake of simplicity, we deduce exporter fixed effects (A_s) and

⁷Elasticities α , β , γ and ε also play an important role in determining the equilibrium solution of our model. However, they are replaced with real world data in our estimations and are therefore only discussed in chapter 5.

⁸This section is based on Bröcker (1998a).

importer fixed effects (B_r) from equation (3.18):

$$A_s = S_s p_s^{-\sigma} \quad (3.19)$$

$$B_r = \frac{D_r}{\sum_s S_s (p_s \tau_{sr})^{-\sigma}}. \quad (3.20)$$

This leaves us with a gravity equation of the form

$$X_{sr} = A_s B_r (\tau_{sr})^{-\sigma}. \quad (3.21)$$

Using equations (3.11) and (3.12) yields the empirical specification of the trade equation. Defining \mathcal{I}_l as the region set of country l , we can state the calibration procedure as follows:

$$X_{sr} = A_s B_r \exp[-\sigma \zeta (g_{sr})^\kappa] z_{lk}^{-\sigma} \quad (3.22)$$

$$\sum_r X_{sr} = S_s \quad (3.23)$$

$$\sum_s X_{sr} = D_r \quad (3.24)$$

$$\sum_{s \in \mathcal{I}_l, r \in \mathcal{I}_k} (X_{sr} + X_{rs}) = T_{lk} + T_{kl}, \quad k \neq l, \quad (3.25)$$

where X_{sr} denote calibrated regional trade flows and T_{lk} are the observed international trade flows.

The restrictions (3.23) to (3.25) ensure that

- the sum of exports from a region s to all regions equals the supply value of tradables in a region s (3.23),
- the sum of imports from all other regions to a region r equals the value of tradables demand in a region r (3.24),
- the calibrated interregional trade flows have to reproduce the respective observed international flows (3.25). In order to assure that estimated trade impediments z_{lk} are unique, we need to assume symmetrical impediments $z_{lk} = z_{kl}$. This is why we calculate the sums $X_{sr} + X_{rs}$ and $T_{lk} + T_{kl}$.

The complete system of equations (3.22)-(3.25) can be solved for A_s , B_r , X_{sr} , and z_{lk} by iterative scaling. In the process S_s and D_r are calculated according to equations (3.7) and (3.8), relying on the aforementioned variables Y_r , N_r , and G_r . The latter two are calculated as follows: National final demand N_k can be calculated using the formula $N_k = Y_k - \sum_{l \neq k} T_{kl} + \sum_{l \neq k} T_{lk}$. Hence, we subtract exports from

country k to all other countries $l \neq k$ from the national GDP in k and add imports from all other countries $l \neq k$ to country k . Regional final demand N_r is calculated by disaggregating the national final demand using regional GDP shares as weights. Transfers G_k are determined by the difference between Y_k and N_k , i.e. $G_k = Y_k - N_k$. Hence, $G_k > 0$ implies a trade surplus. Transfers are also distributed among regions in proportion to their GDP shares.

Once A_s and B_r are given, we can calculate p_s and q_r according to equations (3.19) and (3.10). Numéraire and scaling parameter μ from equation (3.10) are chosen such that prices p_r and q_r have an average of unity across regions.

As we will see in the subsequent section, we incorporate this calibration procedure and its outcomes into our estimation approach of the factor price equation. On the one hand, this enables us to estimate values for σ , ζ , and κ . On the other hand, we can estimate the influence of regional accessibility on regional factor price levels.

3.2.2 The Factor Price Equation

So far we assumed that factor service is produced by a single homogenous factor. For later empirical analysis, we now change this assumption and assume that factor services consist of two components for immobile and mobile factors. We consider factors such as capital as mobile, whereas factors like labor are immobile. The latter is due to the many national borders in Europe which impose high barriers to international migration. Even though European citizens are free to make their residential choice within the EU, language barriers and cultural differences still hamper migration in Europe. An argument in favor of assuming labor as being mobile would be that, after all, we observe large commuter flows in some European regions, at least within national borders. Nevertheless, assuming workers being regionally immobile remains to be realistic because wage differences are not too high among neighboring and nearby regions. This, in turn, means that the over all migration incentive for workers is rather low for short distances.

Formally, we model factor prices for mobile and immobile factors as follows:

$$w_r = \omega_r^\lambda \varphi^{1-\lambda}, \quad (3.26)$$

where ω_r denotes the regional price for immobile factors in region r , φ denotes the regional price for mobile factors, and λ is the income share of labor. We have to separate these two factor service components in order to isolate the immobile factors. Therefore, we fix the prices of the mobile factors $\varphi = 1$, since mobility equilibrates prices over regions in the long run. This leads to $w_r = \omega_r^\lambda$. The latter

modifies equation (3.3) to

$$\omega_r = m_r^{\frac{1}{\lambda\alpha}} p_r^{\frac{1-\beta}{\lambda\alpha}} q_r^{-\frac{\gamma}{\lambda\alpha}}. \quad (3.27)$$

Taking logs of equation (3.27) gives us

$$\log(\omega_r) = \frac{1}{\lambda\alpha} \log(m_r) + \frac{1-\beta}{\lambda\alpha} \log(p_r) - \frac{\gamma}{\lambda\alpha} \log(q_r). \quad (3.28)$$

For the sake of simplicity, we define the accessibility variable ACC_r in the following form:

$$ACC_r = \frac{1-\beta}{\lambda\alpha} \log(p_r) - \frac{\gamma}{\lambda\alpha} \log(q_r). \quad (3.29)$$

Plausibly, in equation (3.28) the factor return is increasing in the accessibility measure. Recall that p_r and q_r are calculated using the calibrated transport costs τ_{sr} and therefore depend nonlinearly on parameters σ , ζ and κ . Thus, we rewrite equation (3.28) as

$$\log(\omega_r) = \frac{1}{\lambda\alpha} \log(m_r) + ACC_r(\sigma, \zeta, \kappa). \quad (3.30)$$

As the productivity m_r is not observable, we assume it to depend on a set of observable variables representing unspecified country effects as well as other influences such as natural endowments and amenities, industry structure etc.⁹ With a log-linear specification we obtain

$$\frac{\log(m_r)}{\lambda\alpha} = X_r \theta, \quad (3.31)$$

where X_r is the vector of observable variables explaining m_r , and θ is a column vector of coefficients to be estimated.

To account for the productivity factors, we rewrite equation (3.30) stacking notation over r as

$$\log(\omega) = \mathbf{X}\theta + \mathbf{ACC}(\sigma, \zeta, \kappa) + \mathbf{u}, \quad (3.32)$$

where \mathbf{X} is the matrix with rows X_r . We also add an additive random disturbance \mathbf{u} . As we will see in chapter 4, this specification allows us to estimate the factor price equation nonlinearly by holding elasticities α , β , γ , and ε as well as λ fixed. In order to adjust for currency effects, we will use PPP-corrected data on GDP when estimating. This assures comparability of the data across countries.

The empirical model in equation (3.32) has the advantage that it builds on all information provided by the underlying theoretical model. We do not need simplifying assumptions on certain variables or elasticities. However, a possible disadvantage is that nonlinear econometric estimations tend to be unstable when applying instru-

⁹See chapter 5.

mentation or spatial correction methods additionally. For instance, this phenomenon is observed in the study of Niebuhr (2004), already reviewed in chapter 2. Therefore, a couple of authors (Mion (2004), Brakman *et al.* (2006)) criticize such nonlinear strategies and prefer to estimate linearized models. We will come back to this issue in chapter 6.

3.3 Chapter Summary

In this chapter we reviewed the *CGEurope Model* designed by Bröcker (1995, 1998a). It serves as the workhorse for our analysis. We developed the empirical specification of the factor price equation. Additionally, we discussed the calibration procedure which is used to calculate regional supply and demand prices. These are based on interregional trade flows and account for national trade impediments as well as distance related transport costs among regions. The procedure is to be incorporated into the estimation of the wage equation. Our exact empirical strategy will be discussed in chapter 4.

Chapter 4

Econometric Theory and Application

In this section we present our econometric approach. We derive cross-section and panel specifications for the model from equation (3.32). In both cases, we illustrate how the estimation of the factor price equation is exactly combined with the described model calibration. Furthermore, we deal with estimation-related problems such as endogeneity of the calibrated exogenous variables, spatial misspecification, omitted variables bias, and the existence of unobservable region-specific effects.

4.1 Cross-Section Specifications

In this section, we turn to the cross-sectional approach. First, we describe our Nonlinear Least Squares (NLS) specification, then we explain the IV and spatial-econometrics approaches which are necessary to account for the econometric problems mentioned above.

4.1.1 Nonlinear Least Squares Approach

In equation (3.32), we stated that we would like to estimate the wage equation depending nonlinearly on the three parameters σ , ζ , and κ . However, computational problems arise in this case as the inclusion of κ prevents our estimation routine from convergence. Hence, we will only estimate the two parameters σ and ζ nonlinearly, and use a predetermined third remaining parameter κ . Bröcker (1998b: 16) estimated a magnitude of $\kappa = 0.58$ in a cross-section for the year 1994. We thus assume something between an exponential and a power form of the transport cost function, yielding a concave curvature. This simplification comes at cost of the functional

flexibility. Since we argued in favor of such a flexibility (see chapter 3), we need to run sensitivity analysis with respect to parameter κ in later regressions.¹

Strategy. When estimating, we keep all elasticities which determine the accessibility measures (α , β , γ , ε , and λ) fixed. Thus, we rewrite (3.32) as

$$\log(\omega) - \mathbf{ACC}(\sigma, \zeta) = \mathbf{X}\theta + \mathbf{u}. \quad (4.1)$$

As we want to adjust to the standard econometric notation, \mathbf{X} in equation (4.1) represents the regressor matrix and not the bilateral trade matrix. On the left hand side, the component \mathbf{ACC} is only nonlinearly dependent on parameters σ and ζ , as we assume parameter $\kappa = 0.58$. For given values of σ and ζ , the coefficient vector θ can be estimated via ordinary least squares (OLS):

$$\hat{\theta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}' (\log(\omega) - \mathbf{ACC}(\sigma, \zeta))). \quad (4.2)$$

We assume that data are normally distributed with $\mathbf{u} \sim N(0, \Omega)$, where Ω is the error covariance matrix. Since cross-sectional specifications of this type usually exhibit heteroscedasticity, we need to apply White's correction mechanism when calculating the standard errors (White (1980)).²

In order to estimate σ and ζ , we need to apply a NLS-optimization routine to equation (4.1). It finds the combination of $\sigma > 1$ and $\zeta > 0$ that minimizes the sum of squared errors (SSE) implied by the OLS-estimator above. The routine has the following steps:

1. Calibrate the *CGEurope-Model* according to equations (3.22) to (3.25), choosing certain starting values for σ and ζ .
2. Calculate regional equilibrium prices \mathbf{p} and \mathbf{q} using the results from the model calibration and obtain the accessibility measures \mathbf{ACC} .
3. Estimate the coefficients θ of the econometric model from equation (4.1) applying an OLS estimator.
4. Calculate the corresponding SSE.

¹Brakman *et al.* (2006) detected estimates of $\kappa = 0.35$. We prefer to use the value of Bröcker (1998b) as he relies on a country set comparable to ours. We leave values like the one from Brakman *et al.* (2006) for sensitivity analysis.

²If errors are iid, we have $\Omega = \sigma^2\mathbf{I}$, where σ^2 is the error variance and \mathbf{I} is the identity matrix. For the case of heteroscedasticity, Ω has to be corrected accordingly.

5. Repeat steps 1. to 4. until the combination of σ and ζ is found that minimizes the SSE.

When altering the values σ and ζ in the described algorithm, we keep the trade flows X_{sr} for the year analyzed fix and let the trade impediments z_{lk} vary (see equations (3.22) to (3.25)).

Calculating standard errors for σ and ζ . We cannot directly calculate standard errors for the optimal values of σ and ζ when estimating equation (4.1) via the described algorithm. Therefore, we need to run an *Artificial Gauss-Newton Regression* (AGNR). This is done by linearizing the nonlinear terms around the NLS-estimates $\hat{\sigma}$ and $\hat{\zeta}$. The respective regression equation reads as

$$\log(\omega) - \mathbf{ACC}(\hat{\sigma}, \hat{\zeta}) = \mathbf{X}\theta + \frac{\partial \mathbf{ACC}}{\partial \sigma} (\sigma - \hat{\sigma}) + \frac{\partial \mathbf{ACC}}{\partial \zeta} (\zeta - \hat{\zeta}) + \mathbf{o}, \quad (4.3)$$

where \mathbf{o} are the AGNR residuals. On the left hand side of equation (4.3), we have the regional wages and the vector of accessibility measures, given the NLS-estimates for $\hat{\sigma}$ and $\hat{\zeta}$. On the right hand side, we have – in addition to the exogenous regressors and the error term – the first derivatives of the accessibility measure with respect to σ and ζ , respectively. The derivatives are calculated using a linear approximation. Results from equation (4.3) are used to calculate the standard errors for σ and ζ .³ If the estimated values $\hat{\sigma}$ and $\hat{\zeta}$ are optimal, the AGNR should yield estimates of $(\sigma - \hat{\sigma})$ and $(\zeta - \hat{\zeta})$ that are zero up to numerical accuracy.

Inference on Accessibility. In order to test for the significance of the accessibility variable, we apply a LR Test ξ_{LR} . Here, we test the null hypothesis H0 that the distance costs play no role in determining regional accessibility. Hence, $\zeta = 0$ while σ floats freely. The alternative hypothesis is that σ and ζ both float freely within the theoretically required ranges. Thus, under H1 distance costs have an impact on regional income differences. The statistic is easily computed using both the log likelihood L of the model under H1 as well as the one under H0. Assuming that the data are normally distributed with $\mathbf{u} \sim N(0, \Omega)$, the log likelihood is

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln |\Omega| - \frac{1}{2} tr(\mathbf{u}'\Omega^{-1}\mathbf{u}), \quad (4.4)$$

where n is the number of observations, \mathbf{u} are the errors from the model in equation (4.1), $|\cdot|$ is the determinant operator, and $tr(\cdot)$ is the trace operator.

³See e.g. Davidson & MacKinnon (2004), chapter 6.5, for methodological details on AGNR.

Given this, the test LR statistic reads as

$$\xi_{LR} = 2 \left[\ln L(\hat{\theta}) - \ln L(\tilde{\theta}) \right], \quad (4.5)$$

where $\hat{\theta}$ is the vector of coefficients in case of H1 and $\tilde{\theta}$ is the equivalent in case of H0 (Verbeek (2004: 172)). Since we rely on two parameters when calibrating the accessibility measure, the statistic is asymptotically χ^2 distributed with 2 degrees of freedom. A high value of ξ_{LR} proves a significant influence of regional accessibility on the income levels and leads us to reject H0.

4.1.2 Instrumental Variables Estimation

We might run into endogeneity problems when estimating equation (4.1). Endogeneity emerges through the fact that we rely on regional GDP measures, both to calculate regional accessibility and to proxy wages ω . Therefore, the accessibility measure might not be statistically independent of the error term. In order to correct this potential deficit, we apply an instrumental variable estimation. We choose the distance to the geographic center of Europe as instrumental variable (**IV**) for **ACC**, as this measure appears to be highly correlated with **ACC**. Here, Luxembourg is defined as the center of Europe.

Strategy. The procedure is a two step Generalized Least Squares (GLS) estimation and is carried out as follows (Wooldridge (2001), chapter 5.1.2). As above, we calibrate our model and calculate prices **p** and **q** to obtain **ACC**. Then, we split the former step 3. into the following two calculations.

- 3a. First, we regress the obtained **ACC** measure on all explanatory variables described above, plus the instrumental variable chosen. Analytically, we have for this first step regression

$$\mathbf{ACC}(\sigma, \zeta) = \mathbf{X}_{IV} \theta_{IV} + \mathbf{v}, \quad (4.6)$$

where $\mathbf{X}_{IV} = (\mathbf{X}, \mathbf{IV})$, $\theta_{IV} = (\theta', \vartheta_{IV})'$, ϑ_{IV} is the scalar coefficient for **IV**, and **v** are the regression residuals.

- 3b. Second, we calculate a fitted accessibility variable $\widehat{\mathbf{ACC}} = \mathbf{X}_{IV} \hat{\theta}_{IV}$. This gives us the modified second step estimation

$$\log(\omega) - \widehat{\mathbf{ACC}}(\sigma, \zeta) = \mathbf{X}\theta + \tilde{\mathbf{u}}, \quad (4.7)$$

where $\tilde{\mathbf{u}}$ are the regression residuals.

Now, we proceed as before.

However, standard errors and inference are based on the recalculated instrumentation residuals

$$\tilde{\mathbf{u}}_{IV} = (\log(\omega) - \mathbf{ACC}(\hat{\sigma}, \hat{\zeta})) - \mathbf{X}\hat{\theta}, \quad (4.8)$$

where $\hat{\theta}$ is the vector of estimated coefficients from equation (4.7). Because simply using residuals from the second step estimation in equation (4.7) would yield biased results (see Verbeek (2004: 146) for details).

Durbin-Wu-Hausman Test. We also check whether we really have to instrument or not. Therefore, we test for exogeneity of \mathbf{ACC} by means of the Durbin-Wu-Hausman Test (DWH – Hausman (1978)). According to Verbeek (2004: 135) or Wooldridge (2001: 119), the null hypothesis of exogeneity of \mathbf{ACC} is rejected if coefficient ϑ in the auxiliary regression

$$\log(\omega) - \mathbf{ACC}(\hat{\sigma}, \hat{\zeta}) = \mathbf{X}\theta + \hat{\mathbf{v}}\vartheta + v \quad (4.9)$$

is significantly different from zero, with $\hat{\mathbf{v}}$ denoting simple regression residuals from equation (4.6) and v being the “new” vector of regression residuals.⁴ In this case, the simple NLS estimator yields biased estimates and we need to apply our instrumentation strategy.

Testing for Overidentifying Restrictions. When instrumenting, we need to test the validity of the instrument choice. This is done by a test of overidentifying restrictions. We apply it in its heteroscedasticity-robust version.

Following Wooldridge (2001: 123), we first regress the chosen instrument on the exogenous regressors \mathbf{X} and the fitted values from the regression in equation (4.6)

$$\mathbf{IV} = \mathbf{X}\theta + \widehat{\mathbf{ACC}} + \mathbf{resid}, \quad (4.10)$$

and keep the new residuals \mathbf{resid} . Then, we regress a vector of ones on the product of \mathbf{resid} and the instrumentation residuals $\tilde{\mathbf{u}}_{IV}$ from equation (4.8). The resulting errors are finally used to calculate the sum of squared errors SSE_{OR} . The latter enters the test statistic. Defining n as the numbers of observations, the test is simply given by

$$\xi_{OR} = n - SSE_{OR}, \quad (4.11)$$

⁴This regression reproduces the IV estimator but also produces an estimate for ϑ .

and asymptotically has a χ^2 distribution with b degrees of freedom, where b denotes the number of overidentifying restrictions. The latter is equal to 1 in our case. We test with the null hypothesis that the chosen instrument is valid, hence that the data are consistent with the model's moment conditions. If we need to reject the null hypothesis – in case the test for overidentifying restrictions is significantly different from zero – the IV specification of the model is rejected.

4.1.3 Spatial Econometrics Problems

Another econometric difficulty stems from the fact that we work with disaggregated regional data. In such a case, regression residuals of neighboring regions tend to be correlated. In order to prevent the applied test statistics from this bias, we need to apply spatial-econometrics tools. This should augment the efficiency of our estimates. In this section, we proceed as follows: First, we define a spatial weighting matrix which describes the spatial pattern of regions, e.g. if regions are neighbors or not. In a second step, we provide test statistics which can be used to detect whether the data exhibit spatial correlation or not. The third step explains how to model spatial dependence econometrically.

Defining an Adequate Spatial Weighting Matrix. We test our model for spatial dependence using a spatial weight matrix \mathbf{W} whose entries mirror the intensity of influence between region pairs. Before normalization, the matrix is symmetric and elements W_{sr} can be metric or binary. We have:

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{pmatrix}. \quad (4.12)$$

By convention, the diagonal elements W_{ss} of the weighting matrix equal zero.⁵

Moreover, results of spatial-econometrics models can depend on the form chosen for W_{sr} (see Florax & Rey (1995)). Additionally, the power of the test statistics can be affected. This forces us to a conservative interpretation of the applied tests. In order to avoid a misleading interpretation of single outcomes, we incorporate several

⁵For statistical and computational reasons, it is common practice to row standardize \mathbf{W} before usage. Row-standardization is the process of normalizing each of the weights in a row such that they sum up to one. However, row-standardizing can change the hypothesized influence of the distance measure. Thus, we need to be careful when economically interpreting results which are based on the parameters of row-standardized weight matrices.

matrices. First, we use a binary neighbors matrix with entries of one for region pairs of neighbors and zero otherwise. Second, we incorporate a metric variant which builds upon the geographic coordinates of a region's centroid. From the scatterplot of coordinates of the regions considered, we create a net of triangles using the *Delaunay-Triangulation Method* (Delaunay (1934)). This method maximizes the smallest inner angle over all triangles. Each triangle consists of three points from the scatterplot of regional coordinates. Note that there must not be a further point in the circumcircle of each triangle. Ergo, each region is assigned to a metric number that mirrors its location in this optimal net. Third, we apply a weight matrix suggested by Bröcker (2005): consider all neighboring regions $s \neq r$ of region r within a distance of 100 minutes travel time. Region pairs with zero distance get a weight of one, whereas region pairs with $g_{sr} \geq 100$ get a weight of zero. For $0 \leq g_{sr} \leq 100$, weights decrease linearly from one to zero with increasing distance. We apply all three measures in order to check the robustness of the outcomes.

Inference on Spatial Dependence. In order to check whether the estimated models suffer from a spatial misspecification either caused by spatial autocorrelation in the error terms (spatial error) or by the dependent variables (spatial lag), we apply the following trinity of tests: a simple Moran-I-Test, a Spatial-Error-Test and a Spatial-Lag-Test.⁶ In the following, we concentrate on the non-instrumented case.

Anselin & Rey (1991) showed that Moran's I has power against both forms of spatial dependence, yet it cannot be used to discriminate between the two. Therefore, we apply the other two test statistics which only test for one kind of spatial dependence.

Moran's I was defined by Moran (1950) as

$$\xi_{MoranI} = \frac{\mathbf{u}'\mathbf{W}\mathbf{u}}{\mathbf{u}'\mathbf{u}}, \quad (4.13)$$

where \mathbf{u} are the regression residuals from equation (4.1). Cliff & Ord (1972) showed that Moran's I asymptotically follows a standard normal distribution. If Moran's I is significantly different from zero, the estimated model generally points to spatial misspecification.

The Spatial-Error-Test checks whether a regression model exhibits spatial autocorrelation in the error term. The test statistic is defined as

⁶For a comparative analysis of these three and various alternative tests see Anselin & Florax (1995).

$$\xi_{SEERR} = \frac{\left(\frac{\mathbf{u}'\mathbf{W}\mathbf{u}}{n}\right)^2}{tr(\mathbf{W}'\mathbf{W} + \mathbf{W}^2)}, \quad (4.14)$$

where n is the number of observations and $tr(\cdot)$ is the trace operator. The test statistic is asymptotically distributed as χ^2 with one degree of freedom (Burrige (1980)).

Finally, the Spatial-Lag-Test detects possible spatial autocorrelation in the dependent variables. Using the NLS results and defining $\mathbf{y} \equiv \log(\omega) - \mathbf{ACC}(\hat{\sigma}, \hat{\zeta})$, we have

$$\xi_{SLAG} = \frac{\left(\frac{\mathbf{u}'\mathbf{W}\mathbf{y}}{n}\right)^2}{\frac{(\mathbf{W}\mathbf{X}\theta)'\mathbf{M}(\mathbf{W}\mathbf{X}\theta)}{n} + tr(\mathbf{W}'\mathbf{W} + \mathbf{W}^2)}, \quad (4.15)$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ denotes the projection matrix and \mathbf{I} is the identity matrix (Anselin (1988a)). The test statistic is asymptotically distributed as χ^2 with one degree of freedom.

Accounting for Spatial Misspecification. Once spatial misspecification is identified, we need to correct for spatial processes in our data. If we ignored these effects and took the least squares estimates for granted, inference on these estimates would be biased.

For the non-instrumented case, we derive the general spatial model specification from equation 4.1. Using $\mathbf{y} \equiv \log(\omega) - \mathbf{ACC}(\sigma, \zeta)$, we can write

$$\mathbf{y} = \mathbf{X}\theta + \iota\mathbf{W}\mathbf{y} + \mathbf{e}, \quad (4.16)$$

$$\mathbf{e} = \varrho\mathbf{W}\mathbf{e} + \nu, \quad (4.17)$$

where \mathbf{e} are the regression residuals, ν is a vector of innovations, ι and ϱ are scalar autoregressive parameters, both less than one in modulus by assumption (see Anselin (1988b): 57ff).

This general model contains a number of special cases.⁷ We consider only one of them, namely the spatial error model (SEM). This is the special case where parameter $\iota = 0$. We concentrate on SEM-models for the reason that the general problem is quite complex. When estimating both kinds of spatial parameters at the same time, results show that the signs of the spatial coefficients are often completely counterintuitive. Thus, we only apply a SEM model, even if the Spatial-Lag-Test is

⁷A complete list is extensively discussed in LeSage (1999), chapter 3.

significant.

Inserting equation (4.17) into equation (4.16), assuming $\iota = 0$, using $\mathbf{e} = \mathbf{y} - \mathbf{X}\theta$, and rearranging reduces the above stated model to a SEM

$$\mathbf{y}^* = \mathbf{X}^*\delta + \nu, \quad (4.18)$$

where $\mathbf{y}^* = \mathbf{y} - \varrho\mathbf{W}\mathbf{y}$, $\mathbf{X}^* = \mathbf{X} - \varrho\mathbf{W}\mathbf{X}$, and δ is the vector of coefficients.

Estimation via Spatial GMM. When applying spatial models, we follow the Feasible Generalized Spatial 2SLS (FGS2SLS) estimator proposed by Kelejian & Prucha (1998, 1999) which belongs to the family of Spatial Generalized Methods of Moments (SGMM). Spatial Maximum Likelihood (SML) techniques are an alternative, but there are two reasons why we prefer to rely on SGMM methods. First, the nonlinear estimation is not stable in the SML case when estimating both parameters σ and ζ . We would have needed to assume a value for one of them in order to guarantee stability. In our eyes, this seems too costly. Second, the SGMM-estimator does not rely on the assumption that the error terms are distributed normally. However, one should not forget to mention a disadvantage of SGMM methods. The outcomes might depend on the moment conditions chosen. Therefore, we need to be careful when interpreting the estimation results.

As shown by Kelejian & Prucha (1999), δ from equation (4.18) cannot be consistently estimated via OLS. The authors propose the following three step estimator.

1. Use OLS to get a starting vector of regression errors.
2. Estimate ϱ by solving the system of moment conditions.
3. Given $\hat{\varrho}$, estimate δ by the FGS2SLS estimator:

$$\tilde{\delta} = (\mathbf{X}^*(\hat{\varrho})'\mathbf{X}^*(\hat{\varrho}))^{-1}\mathbf{X}^*(\hat{\varrho})'\mathbf{y}^*(\hat{\varrho}) \quad (4.19)$$

(see Kelejian & Prucha (1998) for details).

The algorithm described in section 4.1.1 changes to:

1. Calibrate the model according to equations (3.22) to (3.25), choosing certain starting values for σ and ζ .
2. Calculate regional prices \mathbf{p} and \mathbf{q} using the results from the model calibration and obtain the accessibility measures.

- 3a. Compute an initial set of residuals $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\delta}_{OLS}$.
- 3b. Carry out FGS2SLS which yields $\hat{\varrho}$ and $\tilde{\delta}_{FGS2SLS}$.
4. Compute a new set of residuals $\nu^* = \mathbf{y}^* - \mathbf{X}^*\tilde{\delta}_{FGS2SLS}$. Calculate the corresponding SSE which is corrected for spatial processes.
5. Repeat steps 1. to 4. until the combination of σ and ζ is found that minimizes the SSE.

Note, that we do not apply White's standard errors when estimating the spatial models even though these cross-section-specifications might exhibit heteroscedasticity. The reason is that the projection matrix, i.e. the derivatives of the estimated parameters with respect to the residuals, is far more complex than in the NLS case which renders the correction procedure inappropriately difficult. We therefore rely on the normal standard errors.

Inference on Accessibility. In order to test for the significance of the accessibility variable, we now have to take the spatial correction into consideration. Following Anselin (1988b): 182, we first assume $\nu \sim N(0, \Omega)$. Second, an appropriate estimator for the error variance σ^2 is $s^2 = 1/n(\nu^*\nu^*)$. Using this estimator, the spatial log likelihood function can be written as

$$\ln L = -\frac{n}{2}(1 - \ln(\pi)) + \ln |\mathbf{I} - \varrho\mathbf{W}| - \frac{n}{2} \ln \left(\frac{1}{n}(\nu^*\nu^*) \right), \quad (4.20)$$

where $|\cdot|$ is the determinant operator. The test statistic ξ_{LR} itself is identical to the one stated in section 4.1.1. The test statistic is asymptotically χ^2 distributed with 2 degrees of freedom. Large values of the statistic lead us to reject the null hypothesis, meaning we prove a significant impact of the accessibility variable on regional income levels.

Instrumentation. In analogy to the NLS strategy outlined above, the spatial model is used both in a "simple" as well as in an instrumented specification. Instrumentation is carried out as in the NLS case, using a two step procedure. In the procedure described above, we now apply spatial correction mechanisms through our SGMM approach instead of Least Squares methods, i.e. we replace the OLS-estimator θ in step 3b. of the instrumentation strategy, i.e. in equation (4.7), by the SGMM-estimator $\tilde{\delta}$.

4.2 Panel Specifications

Relying exclusively on cross-section specifications causes problems for two reasons. This could lead to an omitted variables bias because our control variables set is not complete in terms of all potential variables explaining regional productivity.

In order to account for this, we apply panel data techniques, too. The time horizon of our data is long enough and we have a complete and consistent data set for each region considered. Panel data models prevent our analysis from the mentioned shortcomings since omitted variables and region-specific effects are gathered in individual intercept terms or random individual effects, respectively.⁸

4.2.1 Nonlinear Least Squares Approach

Turning to panel data models, we extend equation (3.30) by a time dimension $t = 1, \dots, T$:

$$\log(\omega_{rt}) = \frac{1}{\lambda_\omega \alpha} \log(m_{rt}) + ACC_{rt}(\sigma, \zeta). \quad (4.21)$$

We specify $\log(m_{rt})$ linearly with disturbances: $\log(m_{rt})/(\lambda_\omega \alpha) = X_{rt}\theta + u_{rt}$. Hence, all disturbances u_{rt} are assigned to the measures for m_{rt} . Plugging this into equation (4.21) and stacking notation over r and t , we can rewrite the econometric model in equation (4.1) as:

$$\log(\omega) - \mathbf{ACC}(\sigma, \zeta) = \mathbf{X}\theta + \mathbf{u}, \quad (4.22)$$

where $\mathbf{y} \equiv \log(\omega) - \mathbf{ACC}(\sigma, \zeta)$ is now of dimension $nT \times 1$, \mathbf{u} of dimension $nT \times 1$, and \mathbf{X} of dimension $nT \times J$, where J denotes the number of regressors.

For estimation, we prefer a random effects model, since the fixed effects estimator eliminates all time-invariant explaining variables from the model (Verbeek (2004), chapter 10.2). However, an argument in favor of fixed effects panel data models would be the possibility to test the changing regional accessibility given fixed spatial structures. This corresponds well to the questions raised in the beginning of this book. Thus, we will also briefly refer to the fixed effects model.

Following Verbeek (2004), the random effects model exploits two dimensions of the data, between regions and within regions between points in time. Here, individual effects are treated as random variables. They show up as error components in the

⁸Alternatively, one could use difference-in-difference techniques. These were already applied in literature but, at least for the European case, they had very disappointing outcomes which did neither support the model's predictions nor the cross-section evidence (Niebuhr (2004)). Here, we experienced difficulties as well and rather rely on panel methods. In contrast to Mion (2004), we rather rely on a static model because the *CGE Europe Model* itself is static and does not capture endogenous agglomeration processes.

model, represented in stacked notation as

$$\mathbf{y} = \mathbf{X}\theta + \varpi \otimes \mathbf{i}_T + \epsilon, \quad (4.23)$$

where \mathbf{X} is a $nT \times J$ regressor matrix including a constant. The error term is given by $\varpi \otimes \mathbf{i}_T + \epsilon$. It consists of a region specific vector (first term) and a vector of remaining error components (second term). In the first component, \mathbf{i}_T is a $T \times 1$ vector of ones that augments ϖ to panel length using the Kronecker product \otimes .

First, the between dimension is exploited via a regression of individual averages of the regressand, $\bar{y}_r = 1/T \sum_{t=1}^T y_{rt}$, on individual averages of the regressors and a constant, namely $\bar{X}_{rj} = 1/T \sum_{t=1}^T X_{rjt}$. The estimator is given by:

$$\hat{\theta}_{between} = \left(\bar{\mathbf{X}}' \bar{\mathbf{X}} \right)^{-1} \bar{\mathbf{X}}' \bar{\mathbf{y}}, \quad (4.24)$$

where $\bar{\mathbf{y}}$ is a column vector with coordinates \bar{y}_r and $\bar{\mathbf{X}}$ is a matrix with elements \bar{X}_{rj} .

Second, the within dimension of the data is exploited via the fixed effects estimator. One estimates the underlying model in deviations from individual means, that is in stacked notation:

$$\check{\mathbf{y}} = \check{\mathbf{X}}\theta + \check{\epsilon}, \quad (4.25)$$

where $\check{\mathbf{y}} = \mathbf{y} - \bar{\mathbf{y}} \otimes \mathbf{i}_T$ and $\check{\epsilon} = \epsilon - \bar{\epsilon} \otimes \mathbf{i}_T$ are $nT \times 1$ vectors and $\check{\mathbf{X}} = \mathbf{X} - \bar{\mathbf{X}} \otimes \mathbf{i}_T$ is a $nT \times J$ regressor matrix including a constant. Term $\bar{\epsilon}$ is given by the estimated residuals from the regression in (4.24). This yields the within estimator

$$\hat{\theta}_{within} = \left(\check{\mathbf{X}}' \check{\mathbf{X}} \right)^{-1} \check{\mathbf{X}}' \check{\mathbf{y}}. \quad (4.26)$$

Combining both estimators gives the Generalized Least Squares (GLS) estimator. It reads as

$$\hat{\theta}_{RE} = \left(\check{\mathbf{X}}' \check{\mathbf{X}} + \psi T \bar{\mathbf{X}}' \bar{\mathbf{X}} \right)^{-1} \left(\check{\mathbf{X}}' \check{\mathbf{y}} + \psi T \bar{\mathbf{X}}' \bar{\mathbf{y}} \right), \quad (4.27)$$

where ψ is defined as

$$\psi = \frac{\sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + T \sigma_{\varpi}^2}. \quad (4.28)$$

Therefore, the parameter ψ depends on the variance of region specific errors (σ_{ϖ}^2) and of the variance of remaining error components (σ_{ϵ}^2). For $\psi = 0$, the estimator in equation (4.27) reduces to the within estimator. In this case, region specific effects are isolated through the decomposition of the model's constant term, thus each region is virtually treated as a dummy variable.

When calculating the random effects estimator, we proceed in four steps which replace step 3 of the minimization problem described in section 4.1.1:

- 3a. Regress \bar{y} on \bar{X} and a constant via OLS and calculate $\sigma_{\bar{w}}^2$.
- 3b. Regress \check{y} on \check{X} via OLS and calculate $\sigma_{\check{\epsilon}}^2$.
- 3c. Transform all the variables used expressing them as weighted deviations from the individual means, taking into account groupwise heteroscedasticity via usage of ψ . Use transformed variables in OLS in order to obtain $\hat{\theta}_{RE}$.
- 3d. Calculate the random individual intercepts $\hat{w} = \frac{\sigma_{\bar{w}}^2}{T\sigma_{\bar{w}}^2 + \sigma_{\check{\epsilon}}^2}(\mathbf{I}_n \otimes \mathbf{i}'_T)\epsilon$, using the error terms from 3c.

In case of the fixed effects estimator, we eliminate steps 3a, 3b, and 3d of the above stated random effects procedure. The remaining step 3c is adjusted according to $\psi = 0$. Recall that it must not rely on time-invariant regressors.

For both cases, we proceed as described in the NLS case. When applying the random effects and the fixed effects estimator, respectively, we control for time shocks by time dummies in our regression set. Standard errors are calculated in analogy to the cross-section model. Inference is also carried out as for the NLS estimation.

4.2.2 Instrumental Variables Estimation

In section 4.1.2, we chose the time-invariant region's distance to Luxembourg as instrument and applied a 2SLS approach. Since in the panel case the variable to be instrumented, namely ACC_{rt} , varies over time, a time-invariant instrument imposes problems. Therefore, we need to follow a strategy that relies on an instrument which as well adjusts from one time period to another. We opt for instrumenting ACC_{rt} with its own past level. Specifically, we use ACC_{rt-2} , thus lags of two periods.⁹ Besides the different instrument used, we apply the same two-step procedure as in section 4.1.2. Including the lagged instrument in the first step estimation reduces the number of time periods explicitly usable in the second step estimation to six. This leaves us with 7698 observations. In every step described above, the NLS-estimator is replaced by the respective panel estimator.

⁹The natural instrument choice would be a lag of one period. However, time variation of ACC_{rt} is rather low which could yield singularity problems. Therefore, lagging the variable by at least two periods is necessary to really solve the endogeneity problems. See Brakman *et al.* (2004) for a similar strategy in a cross-section setting.

4.2.3 Spatial Econometrics Problems

Baltagi *et al.* (2003) extend spatial LM-Tests to the panel case. In order to test for spatial dependence, we opt for a conditional LM-Test assuming the possible existence of random regional effects (see Baltagi et al (2003: 129f) for details).

Once spatial misspecification is detected, we need to apply spatial econometrics methods to our panel. Again, we use the SGMM estimator, which was modified for panel data in a recent contribution of Kapoor *et al.* (2007). The size of vectors and matrices is now panel augmented (see section 4.2.1).

The first step of the FGS2SLS estimator remains the same. However, in the second step the number of moment estimators needs to be raised from two to three as we now face two error components: one being region-specific (σ_v), and another idiosyncratic one accounting for variations among both the cross-sectional units and the time periods (σ_1). The third step from the cross-section setup generalizes to the following FGLS expression:

$$\tilde{\delta} = \{ \mathbf{X}^*(\hat{\varrho})' [\Omega_\nu^{-1}(\hat{\sigma}_\nu^2, \hat{\sigma}_1^2)] \mathbf{X}^*(\hat{\varrho}) \}^{-1} \mathbf{X}^*(\hat{\varrho})' [\Omega_\nu^{-1}(\hat{\sigma}_\nu^2, \hat{\sigma}_1^2)] \mathbf{y}^*(\hat{\varrho}), \quad (4.29)$$

where Ω_ν is the error covariance matrix (see Kapoor *et al.* (2007) for details). Using these slightly modified three steps, we proceed as in the cross-section case described above, but using the panel augmented FGS2SLS procedure from equation (4.29) instead of the estimator from equation (4.19). Standard errors are calculated in analogy to the spatial cross-section model. Inference is also carried out as in the cross-section case.

4.3 Chapter Summary

In this chapter we explored our econometric approach towards the treatment of the wage equation of the *CGEurope Model*. We described how we combined the model calibration with a nonlinear estimation approach. Then, we proceeded in two steps, namely the presentation of an initial cross-section setup and of a panel setup. In both sections, we tackled the problems of endogeneity and spatial autocorrelation and provided feasible methodological solutions. Concerning spatial econometric models, we opt for SGMM methods rather than SML. The main reason is that the nonlinear estimation is not stable in the SML case when estimating parameters σ , ζ , and ϱ simultaneously. However, there are also concerns using SGMM methods, as results might depend on the moment conditions chosen. Therefore, we need to be careful when interpreting the results.

Chapter 5

Descriptive Data Analysis and Preprocessing

In this chapter we provide a descriptive analysis of the data used subsequently. We use a disaggregated data set and include all NUTS3 regions in the EU-27 into our analysis. This, to our best knowledge, has not yet been done in the literature. For calibration of regional accessibility measures, we also rely on all other European Non-EU regions as well as on the rest of the world.

We proceed with the following five steps. First, we shortly describe the data needed for calibrating our model and for estimating the wage equation. Second, we discuss the regional decomposition of the compiled data set. Third, we take a close look at the data itself in order to understand problems which might arise in later econometric analysis. Fourth, we describe the preprocessing of the data gathered in order to overcome the mentioned problems. Fifth, we conclude with a short chapter summary.

5.1 Data Requirements and Availability

In the calibration procedure described in equations (3.22) to (3.25), A_s , B_r , D_r , S_s , X_{sr} , and z_{lk} are model-endogenous. The same applies to p_r and q_r . Thus, we only need data on T_{lk} , g_{sr} , Y_r , as well as the parameter values of α , β , γ , and ε . For the estimation of the wage equation (3.32), we need data on ω_r and λ . Furthermore, we need data on variables explaining m_r . The latter have to be chosen in accordance with data availability. In literature, there has been a large number of empirical studies that tried to isolate possible alternative determinants of regional income

differences. The authors identify sources from the following non-exhaustive fields:¹

- Natural amenities such as endowments with resources and climate characteristics (Gallup *et al.* (1999), Ellison & Glaeser (1999), Limao & Venables (2001))
- Industry structure (Ellison & Glaeser (1997))
- Education levels (Rauch (1993), Hall & Jones (1999))
- R&D activity and regional knowledge spillover (Jaffe *et al.* (1993), Coe & Helpman (1995), Audretsch & Feldman (1996), Keller (2002) – an overview can be found in Feldman (1999))
- Population density (Ciccone & Hall (1996))
- Participation rates of women in the labor market (Hanson (2005))
- Population age structure (Hanson (2005)).

The time horizon we were able to cover with our respective data set is rather limited. This is due to the fact that regional data for Eastern European countries are only available from 1995 on. Furthermore, the most recent trade data we could find as consolidated data set end in the year 2002. Consequently, our analysis is restricted to the time horizon between 1995 and 2002.

5.2 Regional System

In this section, we take a look at the regional decomposition of our data. Table 5.1 shows the covered countries, broken down to regional entities. The regional system consists of 1372 regions, plus one region accounting for the rest of the world (ROW).² Luxembourg, Liechtenstein, Malta, Cyprus, Iceland, Albania, Macedonia, Moldova, and Turkey are treated on the national level. This is due to their small size and/or to data availability. All other countries are covered on the basis of NUTS3 classification (Eurostat (1999)) or an equivalent.³

¹Rosenthal & Strange (2001) and Head & Mayer (2004) discuss such alternative forces extensively.

²We eliminated all French overseas regions/territories as well as the Spanish Canary Islands from the data set. Generally speaking, we do not consider regions which are – from a geographical perspective – only loosely linked to the rest of their respective country's other regions.

³A detailed description of the current nomenclature is provided by Eurostat on the Internet: http://ec.europa.eu/comm/eurostat/ramon/nuts/home_regions-en.html (last visit March 2008). The last changes became effective in 1999 (reorganization) and 2004 (adding the regions of the new member states).

Cheshire & Hay (1989) discuss the fact that the administratively defined NUTS regions do not mirror regional economies properly. This is especially true in case of disaggregated data, e.g. the NUTS3 system. For instance, in large agglomerations like Hamburg, NUTS separates the economic core from the economic hinterland. Thus, commuter flows are ignored which leads to a measurement bias in economic indicators. Moreover, agglomerations like London or Paris are cut into pieces. Cheshire & Hay (1989) propose an alternative regions concept, so-called Functional Urban Regions (FUR), which eliminates the described problem. However, data are not – or only scarcely – collected for such entities. But, in principle, there are alternatives. For Germany, there exists the concept of *Raumordnungsregionen* which mirror regional economies better. However, we were not able to collect enough data for our analysis that follows this alternative concept. A solution could be not to rely on NUTS3 but on less disaggregated NUTS2 data. The latter exhibit the problems mentioned to a far smaller extent. But as we are interested in an analysis on a disaggregated level, we prefer to stick to the NUTS3 regions and to correct the measurement error before estimating.

In table 5.2 we provide some descriptive statistics for regional area size, nominal GDP levels, and population levels. Having a look at the mean area size, we see that in Scandinavia and Poland regions are quite large in average, whereas for instance German regions tend to be of very small size.

We will use the entire regional system for calibration as we need to rely on links between all European regions to mirror the market interdependencies correctly. For the estimation of the wage equation we only use a subsample of 1283 regions which consists of the EU-27 regions and Norway. Other Western European Non-EU countries like Switzerland, Liechtenstein, and Iceland are excluded due to data restrictions.

5.3 Descriptive Data Analysis

Now we take a closer look at the data itself and its sources. First, we review our national data, such as trade, as well as values for parameters, and elasticities. Then, we turn to the regional data, such as regional distance, GDP, population, or the aforementioned productivity controls.

Table 5.1: Regional Entities

Code	Country	Regions	Entity
1	AT	35	Groups of politische Bezirke
2	BE	35	Arrondissements
2	LU	1	Country
3	GE	441	Kreise and kreisfreie Städte
4	DK	15	Amter
5	ES	50	Provincias
6	FI	20	Maakunnat
7	FR	96	Départements
8	GR	51	Nomoi
9	IE	8	Regional authority regions
10	IT	103	Provincia
11	NL	40	COROP regio's
12	PT	28	Groups of concelhos
13	SE	21	Län
14	UK	133	Groups of unitary authorities
15	BL	28	Oblasti
16	CY	1	Country
17	CZ	14	Kraje
18	EE	5	Groups of maakonnad
19	HU	20	Megyék
20	LT	10	Asprikis
21	LV	6	Groups of rajons
22	MT	2	Malta and Gozo
23	PL	45	Groups of powiaty
24	RO	42	Județe
25	SL	12	Statistical regions
26	SK	8	Kraje
27	CH	26	Kantone
27	LI	1	Country
28	NO	19	Fylker
29	AL	1	Country
30	BA	1	Country
31	BY	6	Oblasti
32	HR	2	Groups of županija
33	IS	1	Country
34	MD	1	Country
35	MK	1	Country
36	RU	28	European oblasti
37	TR	1	Country
38	UA	4	Groups of oblasti
39	YU	4	Republics
Sum	Europe	1372	NUTS 3 equivalent

Table 5.2: Comparability of Regional Entities

Country	Mean area	Std.dev.	Mean GDP	Std.dev.	Mean pop.	Std.dev.
AT	2396.0	(1091.4)	5609.1	(8992.9)	230.3	(258.4)
BE	709.7	(406.6)	5451.8	(7963.5)	237.7	(216.7)
LU	2586.4	(0.0)	18251.0	(0.0)	427.4	(0.0)
DE	809.5	(596.3)	4482.5	(6587.0)	186.1	(183.0)
DK	2873.0	(1808.6)	10719.0	(7218.2)	353.9	(169.2)
ES	9951.0	(5038.6)	10880.0	(17206.0)	765.0	(962.5)
FI	15226.0	(19941.0)	5988.9	(8680.4)	257.9	(263.0)
FR	5666.3	(1923.8)	13805.0	(17176.0)	610.8	(467.0)
GR	2580.9	(1238.6)	2258.4	(6063.5)	209.2	(507.5)
IE	8784.1	(4305.8)	10883.0	(9629.5)	466.9	(269.3)
IT	2925.4	(1750.2)	10589.0	(14627.0)	556.5	(613.2)
NL	847.0	(484.6)	9418.1	(9024.7)	393.8	(296.1)
PT	3171.3	(2208.1)	3646.2	(6427.0)	345.5	(388.6)
SE	19568.0	(22980.0)	10694.0	(13779.0)	422.2	(455.0)
UK	1833.2	(2394.1)	9817.6	(11291.0)	443.8	(351.0)
BL	3961.1	(1523.3)	429.2	(549.3)	292.4	(217.7)
CY	9250.0	(0.0)	8666.2	(0.0)	683.1	(0.0)
CZ	5632.9	(2758.8)	3915.3	(2946.6)	734.2	(316.0)
EE	8485.9	(3035.8)	987.6	(1072.3)	277.6	(165.1)
HU	4651.4	(1837.0)	2348.0	(3357.9)	507.1	(363.0)
LT	6530.0	(1988.0)	995.7	(931.4)	362.3	(253.8)
LV	12918.0	(5907.6)	1259.5	(1411.6)	483.1	(285.8)
ML	157.5	(91.1)	1698.3	(1710.1)	202.1	(214.2)
PL	68471.0	(37022.0)	3426.1	(2935.7)	863.7	(403.9)
RO	5676.0	(1738.3)	864.3	(801.5)	534.1	(286.9)
SL	1689.4	(782.5)	1529.0	(1583.4)	165.7	(131.5)
SK	6129.4	(2487.7)	2482.2	(925.3)	673.1	(79.9)
CH	1524.7	(1881.3)	8166.3	(9640.2)	275.2	(286.4)
LI	160.0	(0.0)	2298.9	(0.0)	32.2	(0.0)
NO	17040.0	(12432.0)	6598.1	(6006.9)	234.1	(123.5)
AL	28748.0	(0.0)	3281.0	(0.0)	3064.1	(0.0)
BA	51129.0	(0.0)	4203.2	(0.0)	3774.0	(0.0)
BY	34600.0	(6799.1)	2131.1	(288.2)	1676.0	(782.5)
HR	28305.0	(3684.0)	9499.2	(3935.2)	2235.7	(926.2)
IS	103000.0	(0.0)	7418.9	(0.0)	276.5	(0.0)
MD	33843.0	(0.0)	1482.9	(0.0)	3647.1	(0.0)
MK	25333.0	(0.0)	3570.2	(0.0)	2014.0	(0.0)
RU	609830.0	(2761400.0)	10631.0	(35707.0)	5188.0	(15167.0)
TR	780580.0	(0.0)	170340.0	(0.0)	65670.0	(0.0)
UA	151680.0	(42825.0)	9508.9	(3197.4)	12407.0	(4039.9)
YU	25587.0	(20808.0)	3119.6	(4511.2)	2656.6	(3251.7)

Source: See text. Area in km^2 , GDP (ϕ 1995-2002) in Mio. Euro, population (ϕ 1995-2002) in 1000.

5.3.1 National Data

Concerning the consolidated international trade data, we rely on a calibrated version of the *CGEurope Model* run by the Institute for Regional Research at the University of Kiel. The respective calibration uses *GTAP version 6* trade time series (Dimaranan (2006)).⁴ Trade consists of agricultural and manufacturing trade and data cover all 39 countries⁵ in our data set plus the ROW region.⁶ Consequently, we use a 40×40 trade matrix for our calculations:

$$T_{lk} = \begin{pmatrix} T_{11} & T_{12} & \dots & T_{1\ 40} \\ T_{21} & T_{22} & \dots & T_{2\ 40} \\ \vdots & \vdots & \ddots & \vdots \\ T_{40\ 1} & T_{40\ 2} & \dots & T_{40\ 40} \end{pmatrix}, \quad (5.1)$$

where T_{lk} denotes the trade flows from country l to k . Trade flows are expressed in terms of current U.S.-\$ in *GTAP* and are, therefore, transformed into current Euro by means of yearly average exchange rates taken from Eurostat's *New Cronos* data base.⁷ When calibrating, the trade diagonal is filled with a proxy for internal trade, namely $T_{ll} = S_l - \sum_{k \neq l} T_{lk}$. Hence, we subtract the exports of country l to all other countries $k \neq l$ from the tradables supply in country l (see Wei (1996)).

There is a number of elasticities in the model that have to be specified properly. We can calibrate elasticities α , β , γ , and λ by the respective expenditure shares on the country level.⁸ For notational convenience, we do not use country indices. Expenditure shares are again obtained from the *GTAP*-based calibrated version of the *CGEurope Model* already mentioned above. As shown in table (5.3), the respective values fluctuate around $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.2$, and $\varepsilon = 0.6$. The latter were already used to parameterize the *CGEurope model* in precedent studies (see e.g. Bröcker (2003)).

5.3.2 Regional Data

Besides data on trade, distance and expenditure shares, we need figures for regional distance and nominal GDP in order to calibrate our model and to calculate regional accessibility measures for all 1373 regions. When estimating the wage equation, we

⁴*GTAP* stands for "Global Trade, Assistance, and Production". This model was developed by the Trade Analysis Program at Purdue University (IL).

⁵The number of countries is 'only' 39 because Luxembourg and Liechtenstein are not treated as separate countries in *GTAP*. Instead, they belong to Belgium and Switzerland, respectively.

⁶Like most comprehensive data bases, the *GTAP Model* does not cover trade in services.

⁷<http://epp.eurostat.ec.europa.eu> (last visit March 2008).

⁸Note that we cannot calibrate elasticities on the regional level due to data restrictions.

Table 5.3: National Parameter Values

	α	β	γ	ε	λ
AT	0.5492	0.2912	0.1596	0.7749	0.4309
BE/LU	0.4999	0.2684	0.2317	0.5367	0.4894
DE	0.5276	0.2891	0.1833	0.6120	0.5050
DK	0.5742	0.2927	0.1331	0.6875	0.4655
ES	0.5667	0.2741	0.1592	0.6326	0.5664
FI	0.5409	0.2686	0.1906	0.5850	0.5929
FR	0.5727	0.2721	0.1552	0.6368	0.4673
GR	0.5151	0.2923	0.1926	0.6028	0.6614
IE	0.4980	0.2793	0.2226	0.5565	0.6320
IT	0.5664	0.2581	0.1756	0.5951	0.4981
NL	0.5080	0.3334	0.1586	0.6776	0.4298
PT	0.4748	0.3126	0.2125	0.5953	0.5305
SE	0.5339	0.2967	0.1695	0.6364	0.3953
UK	0.5381	0.3118	0.1500	0.6752	0.5784
BL	0.4925	0.3045	0.2030	0.6000	0.6455
CY	0.6565	0.2380	0.1055	0.5000	0.5778
CZ	0.3659	0.2852	0.3488	0.4499	0.5433
EE	0.3816	0.3145	0.3039	0.5086	0.6453
HU	0.4599	0.2887	0.2514	0.5345	0.5107
LT	0.4284	0.3162	0.2554	0.5531	0.5917
LV	0.4563	0.3262	0.2175	0.5999	0.6890
ML	0.5757	0.1897	0.2345	0.4000	0.5184
PL	0.4914	0.2532	0.2554	0.4978	0.5591
RO	0.4445	0.2208	0.3347	0.3975	0.5621
SL	0.4071	0.2953	0.2976	0.4980	0.5940
SK	0.3631	0.2756	0.3613	0.4327	0.5480
CH/LI	0.5683	0.2717	0.1601	0.6293	0.5148
NO	0.5929	0.2840	0.1230	0.6978	0.5044
AL	0.6319	0.2179	0.1501	0.5921	0.5605
BA	0.5862	0.2347	0.1790	0.5863	0.4909
BY	0.5036	0.2587	0.2377	0.5212	0.7069
HR	0.5862	0.2347	0.1790	0.5863	0.4792
IS	0.5929	0.2840	0.1230	0.6978	0.5044
MD	0.5036	0.2587	0.2377	0.5212	0.7069
MK	0.5862	0.2347	0.1790	0.5863	0.4909
RU	0.5036	0.2587	0.2377	0.5212	0.5318
TR	0.5799	0.2192	0.2008	0.5219	0.6416
UA	0.5036	0.2587	0.2377	0.5212	0.7069
YU	0.5862	0.2347	0.1790	0.5863	0.4909
ROW	0.5339	0.2818	0.1843	0.6046	0.5528

Source: Institute for Regional Research.

only rely on 1283 regions of our data set, namely the EU-27 and Norway. This choice is made due to data restrictions and because regional entities in Non-EU-countries are hardly comparable to the EU-27 regions in terms of area size and population (see table 5.2 for details). For the estimation of the wage equation we would like to rely on wage data. Since data are not available on a disaggregated level for Europe, we need to choose a proxy. The best proxy would be regional GDP per employee (i.e. see equation (3.4)). Unfortunately, data on persons employed do not exist for our disaggregation level nor for Eastern EU-27-NUTS2 regions. Consequently, we have to proxy wages by GDP per resident and need to think about appropriate corrections. When estimating, we rely on PPP-adjusted regional GDP data which are comparable across regions. If we did not use the PPP-adjusted data, we would not account for exchange rate effects. This would lead to an estimation bias. Thus, we require data on regional GDP, regional population, and a preferably high number of control variables in order to specify the wage equation properly.

We rely on the regional distance data set which was compiled by Bröcker *et al.* (2002). This data cover all 1373 regions. Distance is measured as travel time by car between region centroids. These calculations are based on the network database of IRPUD (2001), which contains data for all major links in Europe, including their specific characteristics of speed limits and likelihood of congestion.

We also need a specification for intraregional distances, which account for the distance to a region's home market. We specify g_{rr} as

$$g_{rr} = \frac{3}{4} \sqrt{\frac{area_r}{\pi}}. \quad (5.2)$$

$Area_r$ denotes the size of region r in km^2 . Data for $area_r$ are obtained from *ESPON* database.⁹ Crozet (2004) remarks that this measure is subject to two forms of bias. First, distances between consumers and producers should be smaller than those implied by the disc approximation because consumers and producers tend to be located in and around cities. Second, our approximation is downward biased since we measure the internal distance 'as the crow flies'. This is contradictory to our interregional approach, where we explicitly rely on transportation networks including congestions costs. Nevertheless, we prefer our simple specification compared to others.

⁹European Spatial Observation Program. For information on the contents of this database, please browse <http://www.espon.eu/> (last visit March 2008).

Table 5.4: Data Sources

Country	Source
EU-27+TR	Eurostat <i>New Cronos</i> http://epp.eurostat.ec.europa.eu
AL	<i>Institute of Statistics</i> http://www.instat.gov.al/
BA	<i>Agency for Statistics</i> http://www.bhas.ba/
BY	<i>Ministry of Statistics and Analysis</i> http://www.belstat.gov.by/homep/en/main.html
CH	<i>Swiss Statistical Office</i> http://www.bfs.admin.ch/bfs/portal/en/index.html
HR	<i>CroStat</i> http://www.hnb.hr
IS	<i>Statistics Iceland</i> http://www.statice.is/
LI	<i>Liechtenstein Department for Economics</i> http://www.llv.li/amtstellen/llv-avw-statistik/
MD	<i>Statistica Moldovei</i> http://www.statistica.md/?lang=en
MK	<i>State Statistical Office</i> http://www.stat.gov.mk/
NO	<i>Statistics Norway</i> http://www.ssb.no/english/
RU	<i>First Independent Rating Agency</i> http://demo.fira.ru/
UA	<i>State Statistics Committee of Ukraine</i> http://www.ukrstat.gov.ua/
YU	<i>Serbia and Montenegro Statistical Office</i> http://www.statserb.sr.gov.yu/

Last visit on websites was in March 2008.

Data sources for EU-27 and Turkey. We exploit Eurostat’s *New Cronos* database to gather data for the EU-27.¹⁰ This gives us a complete data set with respect to regional GDP in PPP-Euro and regional population. With respect to control variables, we can retrieve sectorally disaggregated values for regional Gross Value Added (GVA). However, the sectoral disaggregation is only among three sectors, namely agriculture, industry, and services. Additionally to the data collected from the *New Cronos* database, we gathered regional information on solar irradiation (in kilo Watt hours) and relative urban area¹¹ from the *Joint Research Center* (JRC) of the EU (Suri *et al.* (2005)).¹²

Unfortunately, there are not as many control variables available as we would like to use. For example, we are lacking possible explanatory variables such as education levels of employees, age structure of the population, female participation rate or indicators of R&D activities (such as number of patent applications or R&D investments). Data on the variables mentioned are only (partly) available for NUTS2 regions. This is a serious problem when estimating cross-section models.

GDP time series from 1995 to 2002 are provided by Eurostat for both versions of nomenclature, i.e. the one before 1999 and the one after 1999. However, we have to take care when using other variables. Time series from 1995 to 2002 are not entirely available concerning population and sectoral GVA shares. For instance, population is reported following the “old” nomenclature for years up to 1999, whereas time series following the “new” nomenclature are only provided for the period 1998-2004. The same is true for sectoral GVA shares. Solar irradiation data are also provided using the “new” definition of regional entities. We consolidate our data set on the “old” definition of NUTS nomenclature.

Data Sources for Non-EU Members in Western Europe. For Switzerland, Liechtenstein, Norway, and Iceland, we consult the respective national statistical office. Required data with respect to population and GDP is entirely available. Data on sectoral GVA shares are available for Norway but are lacking for Switzerland, Liechtenstein, and Iceland. Figures of nominal GDP are reported in national currency, hence we transform them into Euro using the yearly average of daily fixing of each national currency in Euro from Eurostat’s *New Cronos* database. PPP-deflators for Norway are taken from Eurostat’s *New Cronos* database. The four countries are also covered by the JRC database on solar irradiation and relative urban area.

¹⁰The main data sources are listed in table 5.4.

¹¹Defined as the share of urban area relative to a region’s total area.

¹²The data can be accessed on the Internet at <http://re.jrc.ec.europa.eu/pvgis/apps/pvreg.php?lang=en&map=europe> (last visit March 2008).

Data Sources for Non-EU Members in Eastern Europe. Data availability is much more scarce concerning Eastern European countries which are not (yet) EU members. However, we only consider this group of countries in the calibration of our model, thus data requirements only comprise nominal regional GDP and regional area. The area size is again taken from *ESPON* data base. Concerning regional GDP, our choice of regional disaggregation comes at cost of comparability and – at least in certain cases – of minor reliability of statistical sources. We partly need to aggregate data to our region system, especially in the case of Ukraine and Croatia. But as we do not need to disaggregate among regions, there are no severe problems arising from these shortcomings. World levels of GDP and population are taken from the World Development Indicators (International Bank for Reconstruction & Development (Various Years)).

Descriptive Analysis. We now take a closer look at the figures of regional GDP per capita, i.e. at how (un)evenly it is distributed over the Europe and how this regional pattern did change over time. Following the principles of *New Trade Theory*, one would expect that the core regions exhibit higher GDP per capita levels because the described linkages act in their favor. The more peripheral a region is, the weaker it should be in terms of GDP levels.

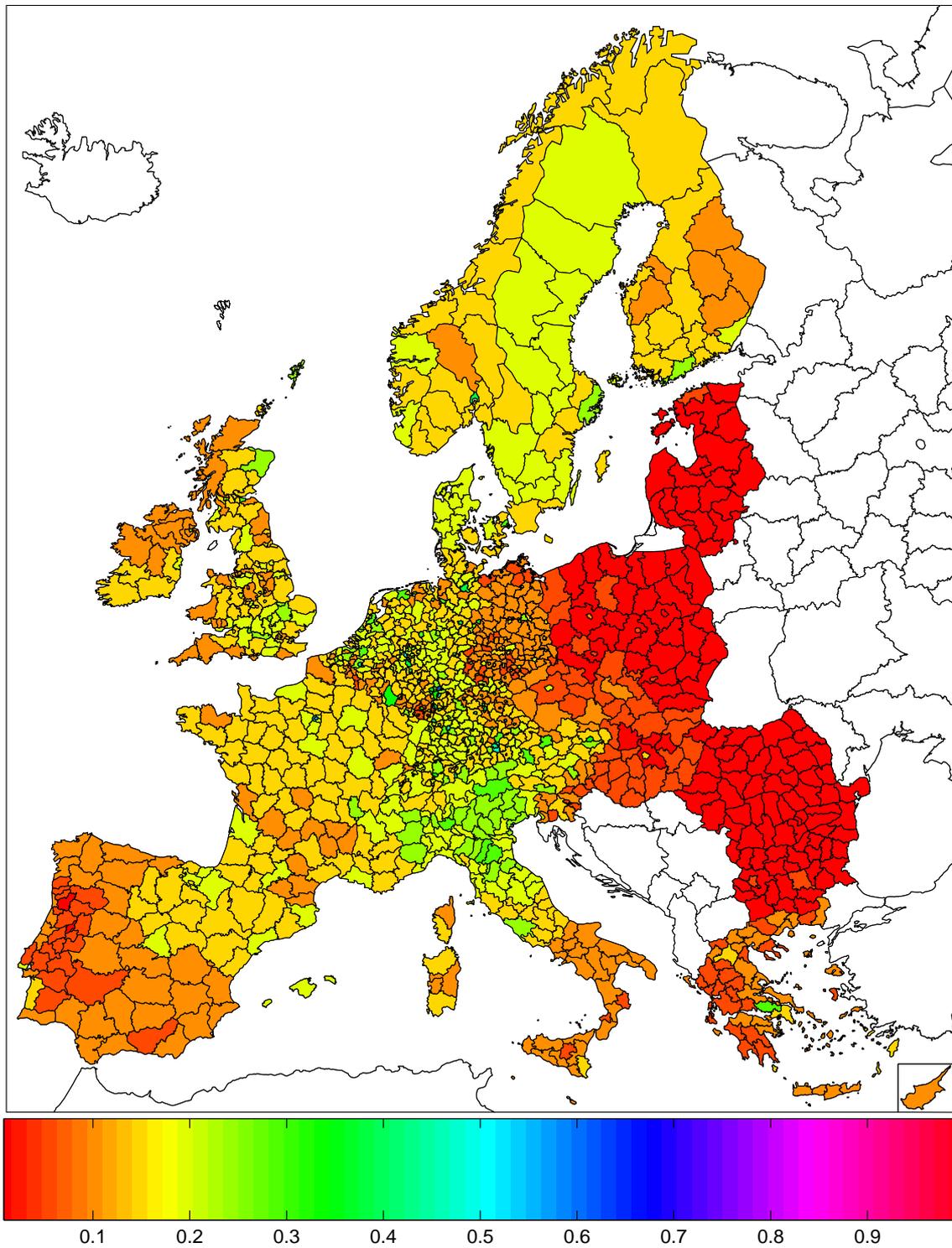
Figure 5.1 depicts the GDP in per capita levels in PPP-Euro in 1995.¹³ Low levels of income are associated with red color. When moving to higher levels, colors change from yellow and green to blue. The highest incomes are marked purple.¹⁴ The region Inner London-West shows the highest level of GDP per capita (72,000 PPP-Euro) whereas the lowest level is assigned to the Bulgarian region Smolyan (2,800 PPP-Euro). Over all, the core European regions exhibit higher income levels than the European periphery, and the single countries' capitals as well as the main economic centers are far richer than the national peripheries. Additionally, one can observe a sharp west-east-decline in GDP per capita levels crossing the EU's eastern border before 2004.

Figure 5.2 depicts the GDP per capita growth from 1995 to 2002. Regions with relatively low income levels in 1995 grew faster than regions with higher initial levels. Moreover, peripheral regions grew faster than central ones. Some regions in Germany even suffer from shrinking GDP per capita levels. We observe the highest GDP per

¹³Graphical representation created with Matlab Arc_Map toolkit (LeSage & Pace (2004)).

¹⁴Note that the different colors are assigned according to the following scheme: Data are transformed to the [0,1]-scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data. Hence, one can only deduce the order of the data from the color bar situated below the figure – not the specific regional value ranges for a certain color or whether values are negative or positive.

Figure 5.1: Regional GDP per Capita Levels (PPP) 1995



Data sources: See text. GDP per resident is expressed in PPP-Euro. Data are transformed to the [0,1]-scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

capita growth in the eastern periphery, whereas levels decreased most in Germany.

As already mentioned, we rather need figures for GDP per employee than the displayed GDP per inhabitant, because the latter do not take into account the commuter flows. Therefore, the centers of agglomerations such as Hamburg exhibit very high GDP per capita levels, whereas its hinterland regions are characterized by considerably lower ratios. This gives a dotted pattern of the GDP distribution across regions – which is especially true for Germany, the BeNeLux-countries and partly the UK, where regional entities are rather small. In order to visualize this problem, we present in figure 5.3 a cutout of figure 5.1, zoomed on the regions in question. We would obtain biased econometric estimates if we took this untransformed ratio as left hand side variable. Hence, we need to smooth out the described mismeasurement.

5.4 Data Preprocessing

To make the available data more suitable for our analysis, some adjustments have to be made. Whereas this preprocessing is only minor in case of trade data as well as parameters and elasticities, they are substantial for our regional wage proxy, i.e. regional GDP per capita.

5.4.1 National Data

The minor adjustments with respect to the national data are the following. The *GTAP Model* does not treat all countries in question as separate entities. Namely, one cannot immediately extract individual information for:

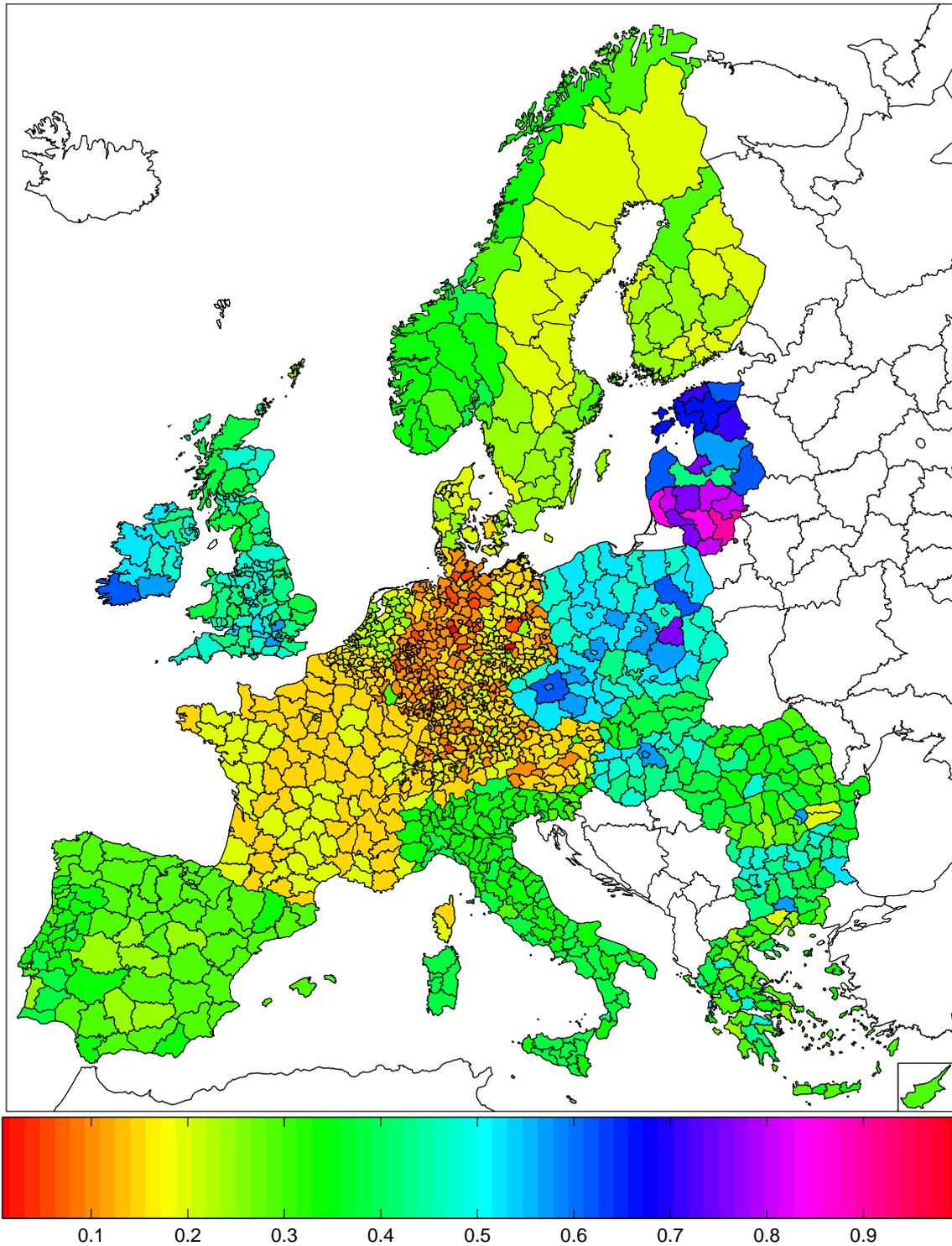
- Norway and Iceland (jointly reported as “rest of EFTA”¹⁵),
- Belarus, Ukraine, and Moldova (among others, jointly covered as “rest of former Soviet Union”),
- Bosnia and Macedonia (among others, jointly mentioned as “rest of Balcan”).

Thus, we need to disaggregate the data from these country groups in order to isolate national values. We do this by using nominal GDP shares as weights.

Concerning parameters and elasticities, modifications are necessary with regard to ε . Since using the original value would give far too high trade estimates for Bulgaria, we set the value equal to the one formerly used, namely $\varepsilon = 0.6$. Comparative adjustments also have to be made for Cyprus ($\varepsilon = 0.5$) and Malta ($\varepsilon = 0.4$).

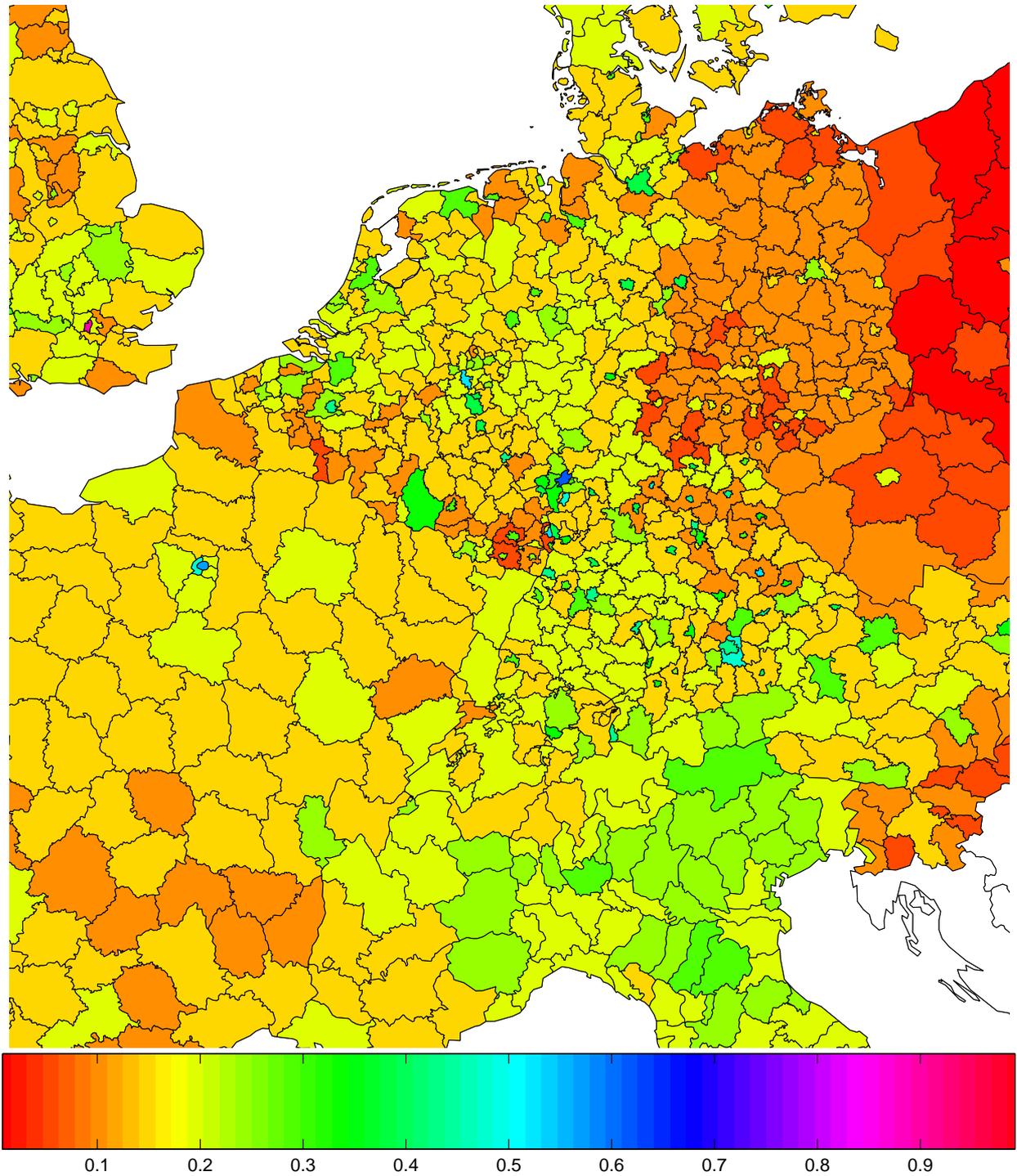
¹⁵European Free Trade Association.

Figure 5.2: Regional GDP per Capita Growth from 1995 to 2002



Data sources: See text. GDP per resident growth rates are expressed in %. Data are transformed to the [0,1]-scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

Figure 5.3: Regional GDP per head (PPP) in Germany (1995)



Data sources: See text. GDP per resident is expressed in PPP-Euro. Data are transformed to the [0,1]-scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

5.4.2 Regional Data

The visualization of the data in tables 5.1 and 5.3 shows that we need to adjust the data for our wage proxy. Regional GDP per head turns out to be biased among strongly agglomerated regions. In order to apply a correction scheme which allows us to keep all degrees of freedom in later regressions, we define a population model which helps us to estimate the population according to place of work. This is done by a gravity model in the spirit of Isard (1954) and Tinbergen (1962). Hence, we use *Newton's Law* to describe the fact that a location exhibits a gravitation force on other locations. This force shrinks with growing distance and depends on a location's relative economic importance, its inner structure, and its location.

We define the estimated population by place of work in region s (L_s) as the total of out-commuters:

$$L_s = \sum_r C_{sr}, \quad (5.3)$$

where C_{sr} is the resident population in r assigned by place of work in s . C_{sr} is assumed to follow a constrained gravity formula,

$$C_{sr} = \frac{P_s \exp(-\chi\sqrt{g_{sr}})}{\sum_t P_t \exp(-\chi\sqrt{g_{sr}})} P_r, \quad (5.4)$$

with P_r denoting the resident population in region r , g_{sr} being the commuting distance between locations s and r in minutes travel time, $\exp(-\chi\sqrt{g_{sr}})$ representing the distance function and χ being the distance decay for commuter flows.

In order to apply this gravity model, we still have to specify the distance decay for commuter flows χ in equation (5.4). We rely on estimates from the literature of commuting behavior: Bröcker (2006) estimates a value of around $\chi = 0.11 \text{min}^{-0.5}$ for two subsamples of European regions, namely Denmark and the German state Schleswig-Holstein. He applies Quasi-Maximum-Likelihood (QML) techniques to the nonlinear model with an additive error.¹⁶ The estimate implies a half-life distance of roughly 6 minutes, which is rather short (see Bröcker (2006: 16f)).¹⁷ We use this estimate as a proxy but not without stating that its importance for the model outcomes is limited. Robustness checks in later empirical analysis only show significant impacts for distance decay values of $\chi > 0.5$.

Taking L_s as the value in the denominator when calculating GDP per capita

¹⁶ P_s and $P_r / (\sum_t P_t \exp(-\chi\sqrt{g_{sr}}))$ are then represented by fixed effects. See Wedderburn (1974) and McCullagh & Nelder (1989) for details on QML estimation.

¹⁷Half-life distance is defined as the distance that reduces the impact to one half of what it would be with zero distance.

measures for region s would already do a better job than simply using resident population figures. But in order to reduce the possible error implied by model (5.3), we apply an additional averaging procedure.

Smoothing Out Mismeasurement. We smooth out mismeasurement by averaging GDP per capita figures with the estimated population by place of work. For any region r which has neighboring regions within a certain radius of minutes travel time, we proceed as follows. We define \mathcal{I}_r as the set of regions within a certain radius of minutes travel time from the center of regions r , including r itself. Then we obtain the corrected GDP per person employed in region r by

$$Y_r^c = \sum_{s \in \mathcal{I}_r} \left(\frac{Y_s}{\sum_{t \in \mathcal{I}_r} L_t} \right). \quad (5.5)$$

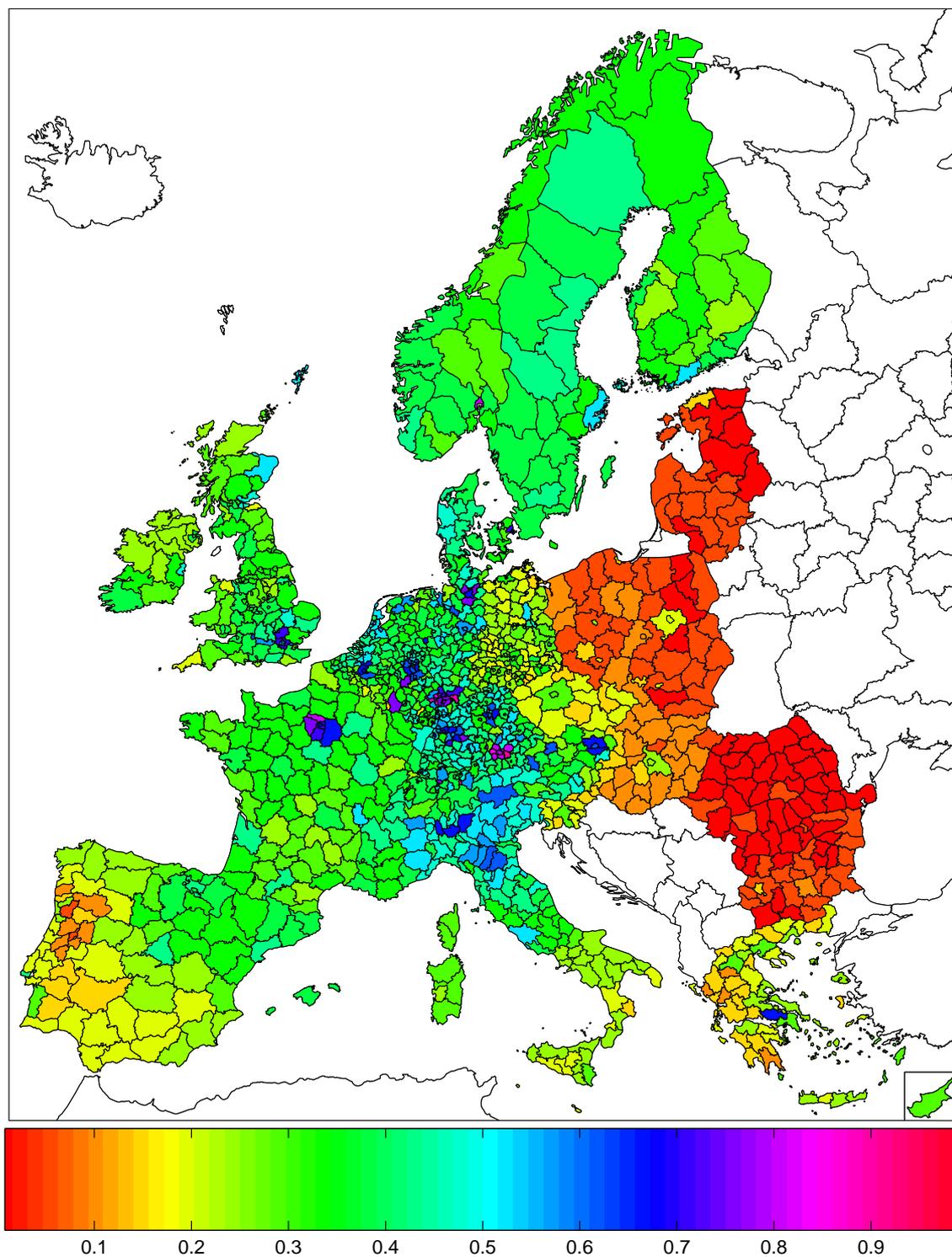
where Y_s is the GDP in region s and L_t is the estimated population assigned by working place in region t . Hence, to each of the regions in the mentioned radius we assign the weighted average income per capita of its surroundings, using the estimated population by place of work as weights. Values for regions which do not have such neighbors remain unchanged.¹⁸

An important decision to make is how big the radius should be, that determines which surrounding regions are included in the averaging procedure. We opt for a radius of 50 minutes travel time, which is based on the literature on regional commuting behavior in European agglomerations and peripheries (Kenworthy & Laube (1999), Schwanen *et al.* (2004)). According to these studies, the average commuting time in agglomerations fluctuates around 30 minutes. However, it is longer in the periphery, where it is more difficult to find work within short distances. Workers therefore do not tend to relocate when finding a job within commuting distance but prefer longer commuting times instead. Commuting distance measured in km had the tendency to grow over the last decades. However, commuting time appears to remain constant over time due to improving transport infrastructure. In the end, we opt for a radius that accounts for average commuting time in agglomerations and in the periphery. Nevertheless, we tried specifications with radii smaller and bigger than 50 minutes but the resulting figures turned out to be less plausible.

The described procedure exhibits the spatial pattern shown in figure (5.4). We see that, especially in Germany, the dotted pattern of the income distribution displayed in figures 5.1 and 5.3 has been eliminated. Again, the Hamburg core and hinterland

¹⁸Note that we have to assure that regions are not aggregated across national borders, since a national border represents a high barrier to commuting.

Figure 5.4: Corrected regional GDP per Capita Levels (PPP) 1995



Source: Corrected GDP per resident is expressed in PPP-Euro, averaged over regions within 50 minutes distance using shares of estimated labor force as weights. Data are transformed to the [0,1]-scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

regions are a good example. Now, we have comparable levels of regional GDP per head in Hamburg city and the surrounding regions. This also means that we have far less outliers in our sample: the maximum (36,000 PPP-Euro) is observed for the German Landkreis Ebersberg in the Munich agglomeration and the minimum, again, for the Bulgarian region Smolyan (still 2,800 PPP-Euro).

5.5 Chapter Summary

In this chapter, we described the regional system underlying our study and the data for calibration and estimation. Particular problems arise when searching for a valid approach to proxy regional wage data, which are lacking for European regions. Plotting the simplest proxy, GDP per resident, shows that it does not exactly display the small scale regional distribution one would expect for regional wages. The latter are supposed to be highest in agglomerations and lowest in the periphery. Moreover, the decrease in wage levels from the center of an agglomeration to its hinterland should be smooth. Data on GDP per resident fits the first picture but fails to fit the second well. It turns out to decrease sharply from the center of an agglomeration to its hinterland. We thus applied an averaging procedure on agglomerated regions which smoothes out the potential mismeasurement and leaves us with a more plausible small scale proxy.

Chapter 6

Estimating the Wage Equation for the Enlarged European Union

In this chapter we present the results of our empirical analysis. Our major aim is to check whether regional market accessibility leads to a significant positive impact on regional income levels. We estimate a set of econometric specifications in order to identify and correct for several econometric problems. Additionally, we perform robustness checks by including productivity controls into our analysis. These are briefly discussed in section 6.1. Similarly to chapter 4, we organize our subsequent analysis in two major sections. At first, we analyze cross-sectional data. Afterwards, we describe the outcomes of our panel specifications. Regarding both alternatives, the panel specifications are principally preferred over the cross-section models. The former should be more robust due to controlling for region-specific effects.

6.1 Preliminaries

When estimating the different econometric models, we will proceed as follows. We estimate the model in equation (3.32), by tracing the productivity m_r back either to an East-West effect or a country effect, plus a set of controls. The first option is to assume that the technology level depends on a region's affiliation to Western or Eastern Europe. Referring to the covered period, the technology level in Eastern Europe was considerably lower compared to the Western counterpart. This fact is displayed by the approximation of a binary dummy variable (called *East*), which is equal to one if the respective region used to be a part of the former *Warsaw Pact* area, and zero otherwise.¹ This dummy variable is expected to exhibit a negative

¹Hence, Germany is a special case because the dummy discriminates between Eastern and Western German regions.

impact on the income level of a region. We call this specification 1 throughout the chapter. A second possibility is to use a set of country dummies (called *Country*). Besides taking into account whether the region belongs to Eastern or Western Europe, this approach might capture any shock on the country level. Hence, we use a binary variable for each country considered. It is equal to one whenever a region belongs to the respective country and zero otherwise. For Germany we introduce two country dummies: one for Western Germany and a second for Eastern Germany. We expect the signs to behave as follows: in general, there should emerge smaller coefficients for Eastern European countries compared to Western European countries. We call this specification 2 throughout the chapter.

In addition, the following controls are used:

- A dummy variable *Capital* which is equal to one if a region comprises the national capital and zero otherwise. Income levels in capital regions are usually higher than in the periphery. For example, in Eastern European regions one observes that e.g. the administration and service sectors are concentrated in the capital region. This fact emerges from their past as centrally planned economies (see e.g. Brühlhart & König (2006)). A comparable pattern is observable in Western European countries, especially when they are centrally organized, i.e. France. We expect this capital dummy to exhibit a positive impact on the regional income level.
- A dummy called *Sea* which equals one, if a region has direct access to the sea, and zero otherwise.² It is to be understood as a measure of tourist attraction. We expect a positive sign for this variable.

Additionally, we implement the following time-variant and time-invariant metric control variables.³

²In the literature, researchers often use a similar measure to identify seaport regions. These benefit from trade since transportation of respective goods – especially over long distances – is largely operated by container shipping. However, this fact is not addressed by our dummy variable. We already use the accessibility measure which itself should account for all these influences.

³The number of included variables is determined via data availability and is not exhaustive. We only report variables that turned out to have a significant statistical impact in later regressions. Other first nature variables, for example the earthquake or volcano risk obtained from the ESPON database, are not mentioned in detail. Recall that the use of the accessibility measure should make the inclusion of geographical characteristics unnecessary.

- Time-variant measures are the industry and agricultural shares of GVA, called *Ind* and *Agrar*.⁴ We expect that *Ind* exhibits a positive impact, whereas *Agrar* has a negative influence on the income level of a region. The rationale behind these presumptions is that industrial regions are usually more developed – not only in terms of economic performance but also with regard to social infrastructure and leisure activities – whereas agricultural regions often are less prosperous concerning the (social) living standards.
- Time-invariant variables are:
 - The solar irradiation in kilo Watt hours (*Kwh*). In this matter, we expect a negative sign because the following regularity usually applies: the higher the solar irradiation, the hotter the climate and the less favorable the general working conditions.
 - The relative urban area (*Urban*) which is defined as a region’s urban area in relation to its total area. This variable should take the “negative side-effects” of agglomerations into account, e.g. pollution, noise, crime or other factors that diminish the capacities of workers and firms in a region. Thus, we expect a negative sign here.

6.2 Cross-Section Results

As already pointed out, estimating cross-section specifications in levels might yield biased results. However, we present a number of these potentially distorted estimates in order to elaborate issues which cannot be accounted for in the panel setup, e.g. that estimated coefficients and parameters might change over time. Moreover, our panel strategy appears to reveal several weaknesses as we will see below. Thus, we proceed as follows. First, we present estimates of the cross-section specifications 1 and 2 for the year 2002 using our NLS and NLS-IV approaches. Second, we report the outcomes for the SGMM and SGMM-IV estimators.⁵ Third, we provide sensitivity analysis results concerning the year of interest and with respect to the shape of the transport cost function.

⁴We have to eliminate one of the three GVA shares since including all three shares would lead to singularity problems when estimating. We choose to exclude the service share. Control regressions showed that including the service share and excluding one of the two other shares never yields significant impacts of the service share. The respective influence might already be captured by the capital dummy.

⁵We use the Spatial Econometrics Toolbox from LeSage (1999) for calculations of the spatial estimators.

Table 6.1: Cross-Section Results: NLS Specification 1

Model	<i>NLS</i>	<i>NLS</i>	<i>NLS</i>	<i>NLS</i>	<i>NLSIV</i>	<i>NLSIV</i>
Proxy	<i>gdph</i>	<i>gdph</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>
Variable	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$
<i>n</i>	1283	1283	1283	1283	1283	1283
σ	12.5930 (0.6937)	14.1384 (0.9338)	11.3376 (0.5419)	11.8676 (0.6512)	9.8910 (0.4671)	10.2497 (0.5343)
ζ	0.0112 (0.0011)	0.0075 (0.0010)	0.0128 (0.0011)	0.0112 (0.0010)	0.0113 (0.0011)	0.0080 (0.0011)
Constant	-3.4132*** (0.0043)	0.3611 (0.3361)	-3.9153*** (0.0080)	-2.5962*** (0.3167)	-3.9150*** (0.0086)	-2.7607*** (0.2995)
East	-0.5174*** (0.0223)	-0.4370*** (0.0188)	-0.5053*** (0.0210)	-0.4383*** (0.0190)	-0.4804*** (0.0236)	-0.4207*** (0.0209)
Capital		0.4459*** (0.0674)		0.3608*** (0.0500)		0.4276*** (0.0520)
Sea		0.0923*** (0.0172)		0.0788*** (0.0163)		0.0867*** (0.0165)
Urban		-0.0764*** (0.0100)		-0.0740*** (0.0087)		-0.0718*** (0.0090)
Kwh		-0.1216*** (0.0466)		-0.2177*** (0.0447)		-0.1900*** (0.0425)
Ind		0.0487* (0.0299)		0.0404* (0.0214)		0.0549** (0.0228)
Agrar		-0.1146*** (0.0118)		-0.0812*** (0.0081)		-0.0782*** (0.0089)
ξ_{LR}	429.2***	241.6***	616.7***	380.2***	394.9***	217.6***
ξ_{MoranI}	0.2278***	0.2684***	0.3783***	0.3435***	0.4990***	0.4144***
ξ_{SERR}	94.8***	373.7***	516.1***	627.9***	937.8***	646.7***
ξ_{SLAG}	2.0428	29.8***	0.1521	0.9859	872.2***	651.6***
ξ_{DWH}					-0.0091	-0.1636***
ξ_{OR} prob					0.3313	0.7632
adj. R^2	0.5909	0.7095	0.5999	0.6965	0.6594	0.7283

White's heteroscedastic standard errors in parenthesis. Standard errors for instrumentation models are based on the adjusted instrumentation error $\tilde{\mathbf{u}}_{IV}$. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, and * denotes significance at the 0.1 level. Tests for weak instruments are not shown in detail but clearly prove the power of the applied instruments. For ξ_{OR} only the t-probabilities are displayed. Spatial tests are performed using a spatial weights matrix with weights decreasing linearly from 1 to 0 while the centroid distance increases from 0 to 100 minutes travel time. Tests are robust to other weighting matrices.

6.2.1 NLS Specifications

In this section, we present the NLS-outcomes for specifications 1 and 2. Because of data restrictions, the econometric analysis is based upon 1283 of the 1373 regions used for calibration. As outlined in chapter 5.4.2, the shortage of regional wage data force us to approximate regional wage levels by regional GDP per residents levels. However, the application of this proxy would lead to distorted estimates. That is the reason why we also evaluate the sensitivity of our estimations with respect to the usage of different wage measures. These are the GDP per resident itself and the one corrected by the estimated work force.⁶

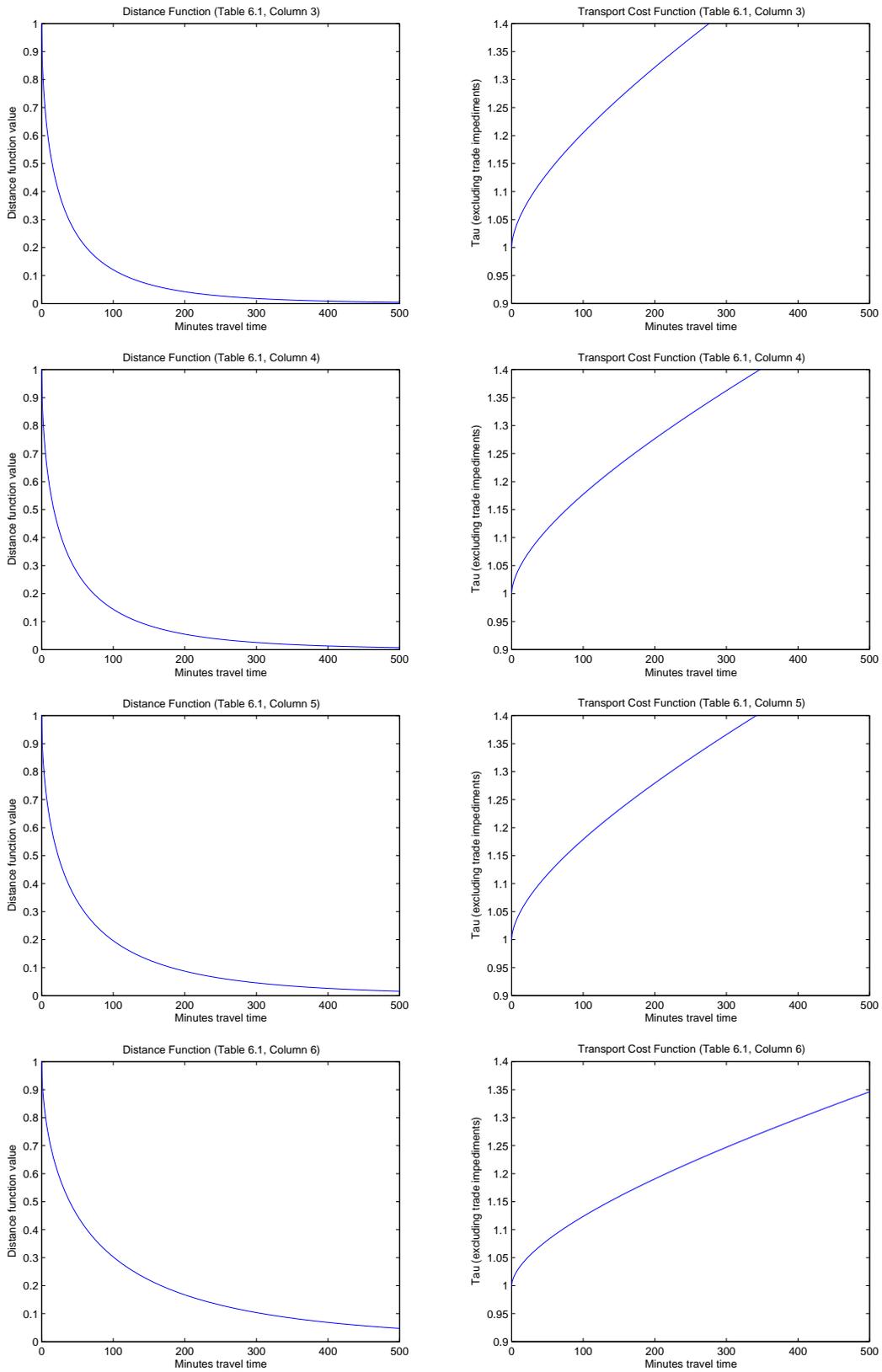
Results for the cross-section specification 1 for the year 2002 are reported in table 6.1. In columns (1) and (2), we use the untransformed regional GDP per head ($gdph$) when calculating the dependent variable; in columns (3) to (6), we use GDP measures ($gdph_{corr}$) that correspond to the described correction procedure. Columns (5) and (6) comprise the NLS-IV approach using $gdph_{corr}$ in the left hand side variable. In all columns, we assume a value $\kappa = 0.5817$. For all setups we estimate two different scenarios: one which only applies dummies to estimating m_r and a second which also includes the aforementioned other productivity controls.

In column (1), we infer from the LR-Test that the null hypothesis of an insignificant impact of accessibility can be rejected. Including wage controls diminishes the size of the effect of regional relative prices on the income distribution. However, the impact remains significant at the 0.01 level. Spatial tests show that there is spatial correlation in the data which seem to pertain in the error term. Therefore, we need to estimate spatial models in order to obtain valid estimates. Subsequently, the results in column (1) should not be taken for granted with respect to the statistical significance and the standard errors, respectively. The same evidence applies to column (2) where we included the productivity controls. In columns (1) and (2), coefficients have the expected signs.

When using the corrected GDP per capita measures in columns (3) and (4), values for the highly significant LR-Test rise. Applying our instrumentation strategy in columns (5) and (6) gives lower LR-levels again. The value of the DWH-Test statistic is significantly different from zero for the model in column (6) which points to endogeneity of the accessibility variable and the necessity to instrument. Contrarily, following the DWH-Test there is no need to instrument in case of the model in column (5). However, the test for overidentifying restrictions is insignificant for both models pointing at a valid instrumentation strategy. Instruments are never

⁶See chapter 5 for details.

Figure 6.1: Distance & Transport Cost Functions (Cross-Section Specification 1)



weak as they exhibit highly significant coefficients in the first step estimation of our instrumentation approach (not reported in detail). Surprisingly, the spatial tests increase in values when applying the instrumentation strategy and the Spatial-Lag-Test becomes highly significant, too. This outcome is robust to usage of other spatial weighting matrices such as binary contiguity matrices or the aforementioned metric variant (results are not shown in detail here).

The model fit is driven up considerably when including the productivity controls in the estimation. It then fluctuates around $R^2 = 0.7$. The R^2 , the coefficients and the standard errors vary slightly across columns. As it should, the instrumentation strategy yields higher standard errors compared to the simple NLS strategy. Remarkably, the *Sea* dummy exhibits a significant impact, even though the model relies on the accessibility measure. Relying on the respective coefficients in column (6) and exponentiating, we can infer the two following exemplary conclusions: the income level in the former *Warsaw Pact* regions is only 65% of the level in Western European regions. And capital regions' income is 1.5 times higher than the one in other regions. In conclusion, results for specification 1 indicate, that NLS-IV estimates in columns (5) and (6) should be preferred over the simple NLS estimates.

The nonlinearly estimated parameters σ and ζ fluctuate around $\sigma = 11$ and $\zeta = 0.01$ across columns. Figure 6.1 depicts the implied distance and transport cost functions. We can see that the shape of these functions is in line with the expectations and that the implied distance decay is plausible. For instance, the transport cost markup deduced from column (6) amounts to roughly 30% over a distance of 400 minutes travel time. We point out that these results are robust to an increase of the aggregation radius even to an implausibly large radius of 100 minutes travel time. The same applies to a variation of the assumed distance decay χ . Here, coefficients do not change remarkably up to a value of $\chi = 0.5$.

Table 6.2 displays the results for the cross-section specification 2. Inference on accessibility shows that the LR-Test decreases considerably in absolute terms for this specification, but remains statistically significant at the 0.01 level. When we include the productivity controls, the absolute value of the LR-test decreases even more but still indicates a significant impact of **ACC**. Hence, usage of country dummies seems to imply a less pronounced impact of accessibility on the regional income level. Because we have included country dummies in the current case, the accessibility measure itself seems to explain the accessibility difference on the national level well, but less within countries. Usage of country dummies increases the model fit to roughly 0.8.

Relatively seen to the other regressors – whose coefficients are often close to zero

Table 6.2: Cross-Section Results: NLS Specification 2

Model	<i>NLS</i>	<i>NLS</i>	<i>NLS</i>	<i>NLS</i>	<i>NLSIV</i>	<i>NLSIV</i>
Proxy	<i>gdph</i>	<i>gdph</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>
variable	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$
<i>n</i>	1283	1283	1283	1283	1283	1283
σ	15.2843 (1.3665)	48.4259 (12.6911)	13.2201 (1.0009)	20.5806 (2.5764)	13.2288 (3.4803)	12.0608 (3.5303)
ζ	0.0160 (0.0011)	0.0060 (0.0010)	0.0170 (0.0013)	0.0110 (0.0012)	0.0078 (0.0008)	0.0051 (0.0110)
Capital		0.3905*** (0.0637)		0.3198*** (0.0526)		0.3384*** (0.0525)
Sea		-0.0028 (0.0763)		-0.0056 (0.0147)		-0.0003 (0.0151)
Urban		-0.0004 (0.0115)		0.0019 (0.0091)		0.0096 (0.0093)
Kwh		-0.0183 (0.0964)		-0.0205 (0.0918)		-0.0148 (0.0957)
Ind		0.0851*** (0.0260)		0.0612*** (0.0213)		0.0557*** (0.0220)
Agrar		-0.0880*** (0.0123)		-0.0461*** (0.0070)		-0.0508*** (0.0071)
Country	yes	yes	yes	yes	yes	yes
ξ_{LR}	209.7***	28.3***	332.1***	114.5***	100.3***	110.2***
ξ_{MoranI}	0.0335***	0.1395***	0.2194***	0.2371***	0.2815***	0.2704***
ξ_{SERR}	12.2***	293.2***	614.7***	800.6***	298.5***	275.4***
ξ_{SLAG}	0.4785	22.51***	10.06***	0.8111	166.3***	256.4***
ξ_{DWH}					-0.2982**	-0.1991***
ξ_{OR} prob					0.9945	0.9995
adj. R^2	0.7276	0.8041	0.8004	0.8284	0.7702	0.8194

White's heteroscedastic standard errors in parenthesis. Standard errors for instrumentation models are based on the adjusted instrumentation error $\tilde{\mathbf{u}}_{IV}$. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level and * denotes significance at the 0.1 level. Tests for weak instruments are not shown in detail but clearly prove the power of the applied instruments. For ξ_{OR} only the t-probabilities are displayed. Spatial tests are performed using a spatial weights matrix with weights decreasing linearly from 1 to 0 while the centroid distance increases from 0 to 100 minutes travel time. Country dummy values are not reported. The use of a full set of country dummies makes the use of an intercept term as well as *East* unnecessary.

– country effects are highly significant and play an important role in explaining the regional income differences.⁷ Country effects have a standard deviation ranging from 0.39 (mean= -4.1) for the logarithmic model in column (6) to 0.55 (mean= -3.6) for column (5). These standard deviations imply considerably large productivity differentials. A difference of 0.39 between country effects corresponds to a productivity ratio of $\exp(0.39) = 1.48$, *ceteris paribus*. For a difference of 0.55 the ratio is 1.73.

In principle, most signs of the linear model coefficients remain the same. Only *Urban* turns from a negative to a positive sign in columns (4) and (6), whereas *Sea* changes to a negative one across all models. Hence, we can conjecture that regions with direct sea access seem to cluster in certain countries and regions without in others. Such a fact is already accounted for by the country dummies. Therefore, it is not surprising that, contrary to specification 1, *Sea* does not exhibit significant statistical impacts. Other first nature controls, such as *Urban* and *Kwh*, neither have a significant impact. Finally, the DWH-Test statistic confirms the endogeneity of accessibility. The insignificant test for overidentifying restrictions confirms the null hypothesis of a suitable IV setup. In every model estimated, spatial tests indicate that spatial econometric tools should be applied for specification 2 as well.

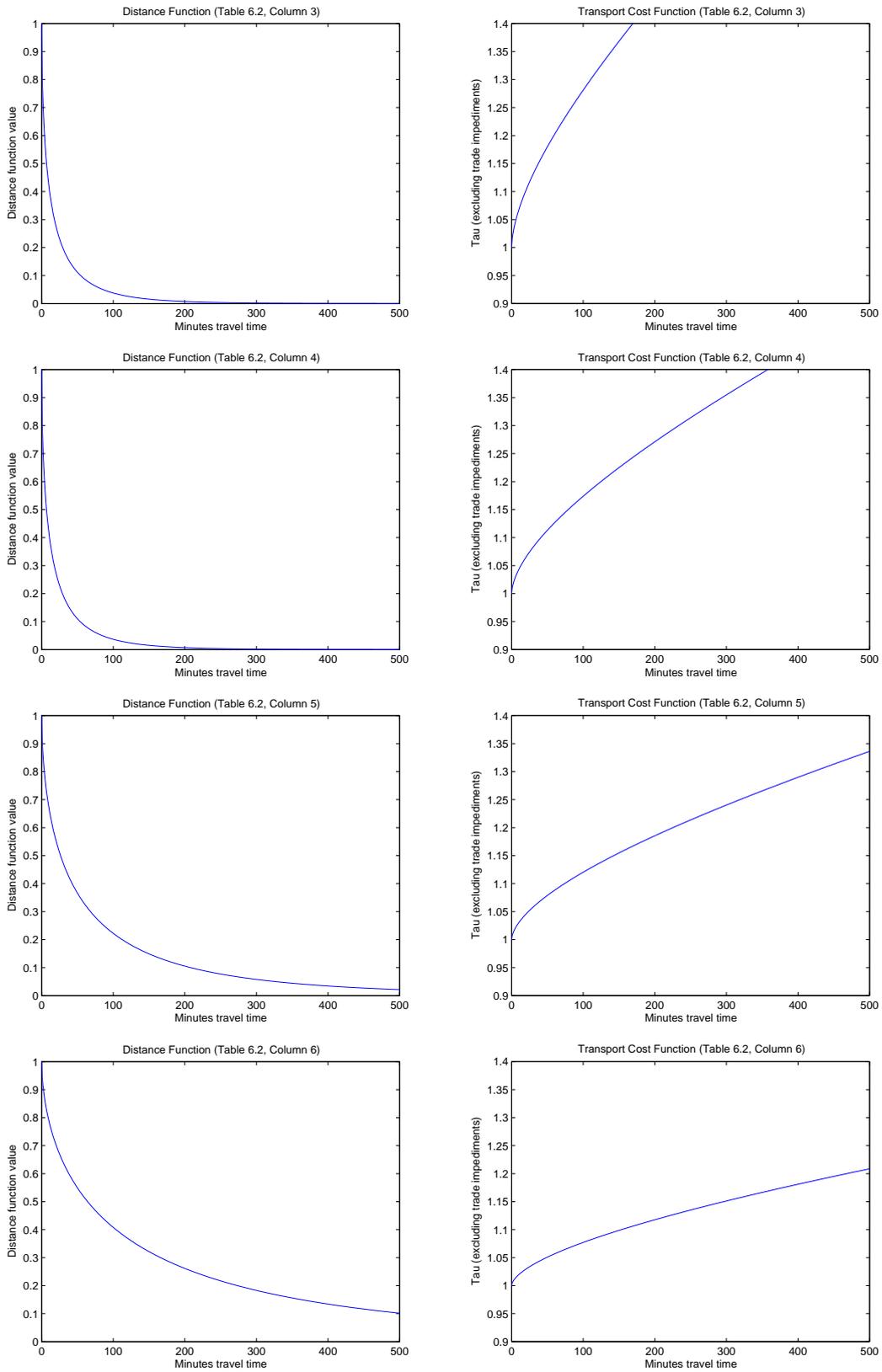
Compared to specification 1, estimates for σ are slightly higher when only relying on country dummies to proxy m_r . However, estimates for ζ are similar in magnitude. Estimates for σ become more uncertain since standard errors rise compared to specification 1. When including controls, estimates for σ are higher than for the models presented above, especially when using *gdph* on the left hand side. In case of our instrumentation strategy, this effect is levelled out. Figure 6.2 shows the respective distance and transport cost functions. It appears that the distance decay is comparable to specification 1 as it should be.

6.2.2 Spatial Econometrics Models

So far, all models exhibit spatial autocorrelation. This has to be corrected for in order to come to unbiased standard errors. We do not estimate a general spatial model because it renders the estimation too complex. Thus, we concentrate on spatial error models. These should correct for the major drivers of the spatial distortions because in a number of specifications tested, the latter seem to pertain in the error term rather than appearing in the regressands. The corresponding results are listed in table 6.3 where we report spatial estimates for specification 1 in columns (1) to (3), and for specification 2 in columns (4) to (6). We only include estimations which rely

⁷Coefficients are not reported in detail.

Figure 6.2: Distance & Transport Cost Functions (Cross-Section Specification 2)



on the corrected GDP per capita measure.

The LR-statistic for the spatial error model reveals, that the impact of the accessibility measure drives down considerably even though it remains significant at the 0.01 level. The spatial autocorrelation coefficient ϱ is always highly significant and the model fit slightly increases with respect to the NLS models – this outcome is again robust to the usage of different weighting matrices.⁸ Hence, in the cross-section case spatial models seem to be much more appropriate for our estimation approach than simple NLS specifications.

Across columns the linear coefficients do not change signs or become less significant compared to precedent models. However, they change their size to a certain extent. For example, the coefficient for *East* falls from -0.7 in column (1) to -1.2 in column (3) when instrumenting which shows that the effect of belonging to the group of regions from former socialist countries now has a more prominent impact. When instrumenting, standard errors increase as they should. The test for overidentifying restrictions confirms our instrumentation strategy.

In total, the σ -values increase for non-instrumentation models compared to the simple NLS estimates. They now range between $15 < \sigma < 27$ which implies a higher substitutability of manufacturing varieties compared to the former models. The ζ -estimates seem to remain on a constant level compared to the respective NLS columns in tables 6.1 and 6.2. Figure 6.3 depicts the shapes of the transport cost and distance functions implied by the spatial models. The transport cost functions slightly flatten referred to the non-spatial models (except for column (5)). Regarding the distance function, the outcome for the distance decay is not unambiguous. For column (2) the function clearly flattens whereas it steepens in case of the model in column (5). The results for the other columns do not change the picture much compared to the models estimated before.

The implications of what we have described so far are visualized in figure 6.4 and 6.5. Here, we depicted the calibrated accessibility measures for 2002. The parameter values of σ and ζ are chosen from simple NLS estimations of specification 1 ($\sigma = 12$ and $\zeta = 0.01$) and 2 ($\sigma = 20.5$ and $\zeta = 0.01$).⁹ We observe a plausible spatial distribution of accessibility measures, identifying the central European regions in light color as those with the best accessibility. Eastern plus distant regions are marked in red and dark colors and are characterized by the worst accessibility, hence the highest demand prices and the lowest supply prices. Better accessibility levels

⁸Respective results are not given in detail here.

⁹We have a look on landscapes implied by parameter estimates comparable to the NLS-IV results below.

Table 6.3: Cross-Section Results: Spatial Models

Model	<i>SGMM</i>	<i>SGMM</i>	<i>SGMMIV</i>	<i>SGMM</i>	<i>SGMM</i>	<i>SGMMIV</i>
Proxy	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>
Variable	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$
<i>n</i>	1283	1283	1283	1283	1283	1283
σ	15.0919 (0.9109)	17.7263 (1.3606)	10.8474 (0.8256)	16.3671 (1.4440)	26.8977 (4.1598)	12.0174 (4.8713)
ζ	0.0135 (0.0011)	0.0106 (0.0010)	0.0056 (0.0009)	0.0150 (0.0010)	0.0094 (0.0011)	0.0046 (0.0009)
Constant	-3.9075*** (0.0134)	-1.6611*** (0.3502)	3.1535*** (0.4573)			
East	-0.6936*** (0.0236)	-0.6609*** (0.0230)	-1.2474*** (0.0310)			
Capital		0.2612*** (0.0386)	0.2583*** (0.0452)		0.2901*** (0.0358)	0.3061*** (0.0369)
Sea		0.0415*** (0.0174)	0.0841*** (0.0220)		-0.0106 (0.0162)	0.0033 (0.0169)
Urban		-0.0307*** (0.0080)	-0.0421*** (0.0102)		0.0115 (0.0083)	0.0128 (0.0086)
Kwh		-0.3404*** (0.0499)	-1.0266*** (0.0650)		-0.0801 (0.1119)	-0.0441 (0.1155)
Ind		0.0497** (0.0182)	0.0354 (0.0222)		0.0725*** (0.0171)	0.0631*** (0.0176)
Agrar		-0.0623*** (0.0057)	-0.0731*** (0.0069)		-0.0417*** (0.0054)	-0.0453*** (0.0056)
ϱ	0.5887***	0.5835***	0.7089***	0.3978***	0.4163***	0.4176***
Country	no	no	no	yes	yes	yes
ξ_{LR}	80.5***	87.1***	61.4***	40.2***	32.6***	9.7***
ξ_{DWH}			-16.51***			5.8964
ξ_{OR} prob			0.1905			0.4671
adj. R^2	0.7631	0.8079	0.7828	0.8231	0.8131	0.9252

Standard errors (based on the spatially corrected regression error) in parenthesis. Standard errors for instrumentation models are additionally based on the adjusted instrumentation error. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level and * denotes significance at the 0.1 level. Calculations are performed using a spatial weights matrix with weights decreasing linearly from 1 to 0 while the centroid distance increases from 0 to 100 minutes travel time. Tests for weak instruments are not shown in detail but clearly prove the power of the applied instruments. For ξ_{OR} only the t-probabilities are displayed. The LR-Test is based on a spatial log-likelihood function. Country dummy values are not reported in detail. The use of a full set of country dummies makes the use of an intercept term as well as *East* unnecessary.

Figure 6.3: Distance & Transport Cost Functions (Spatial Cross-Section Models)

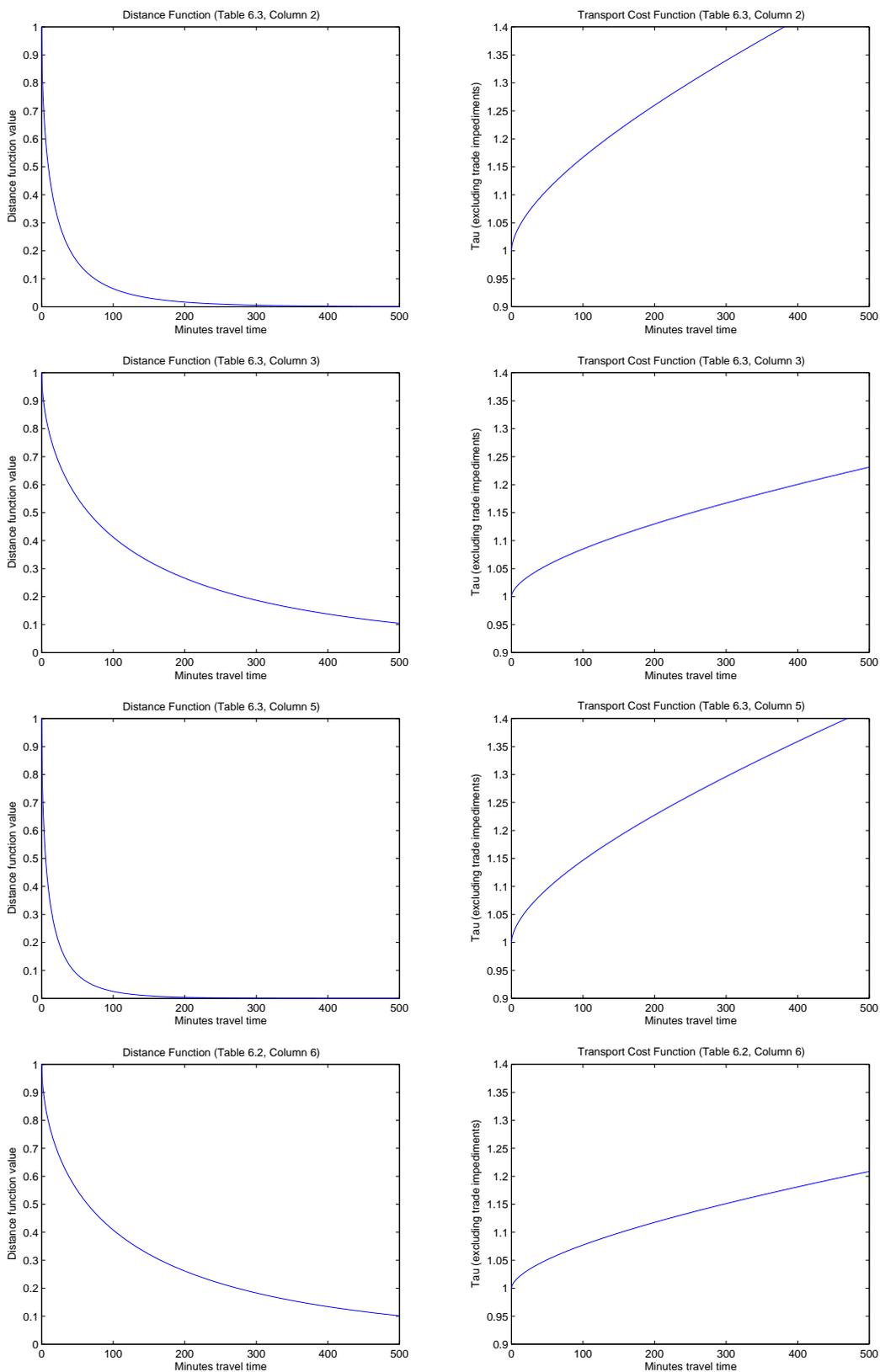
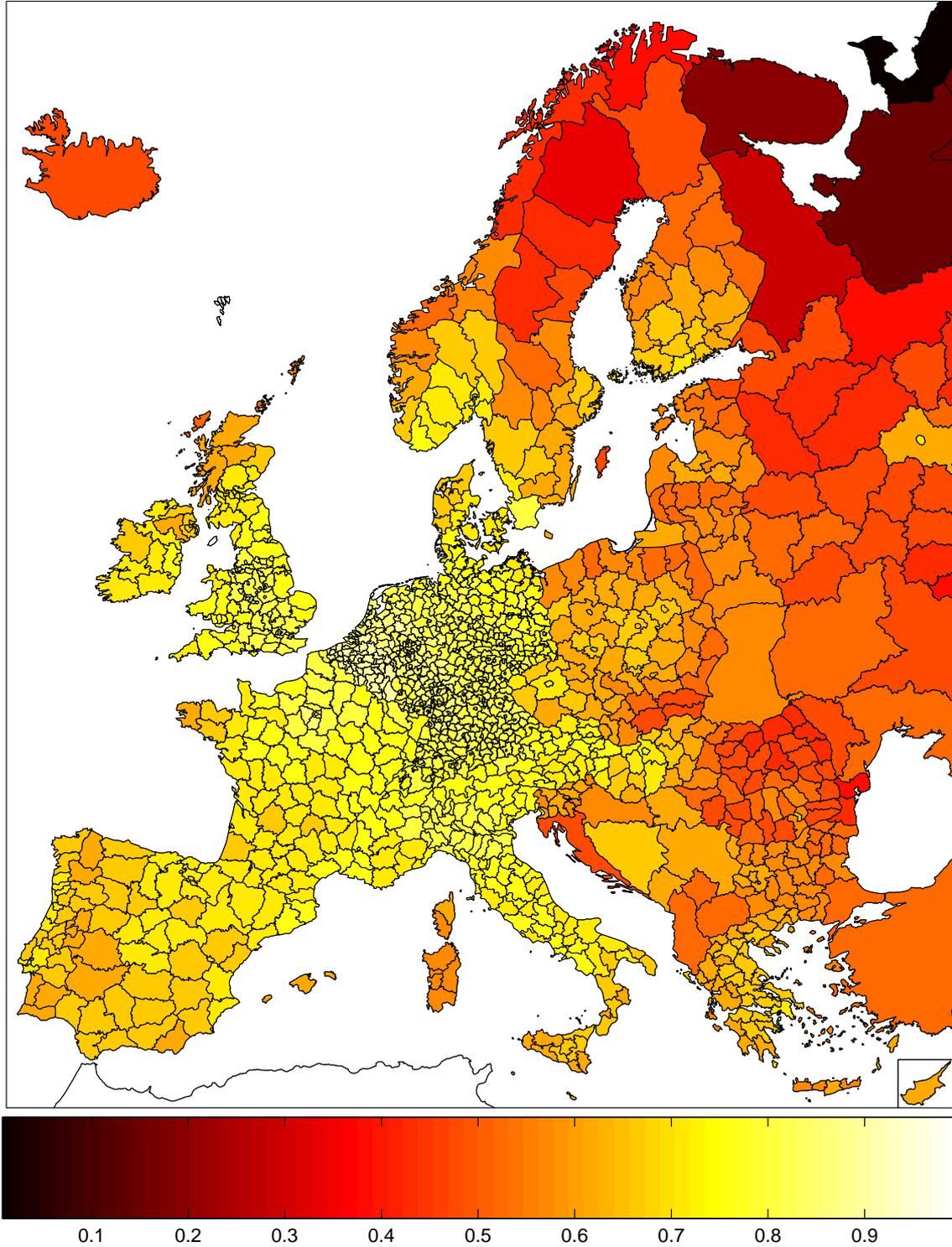
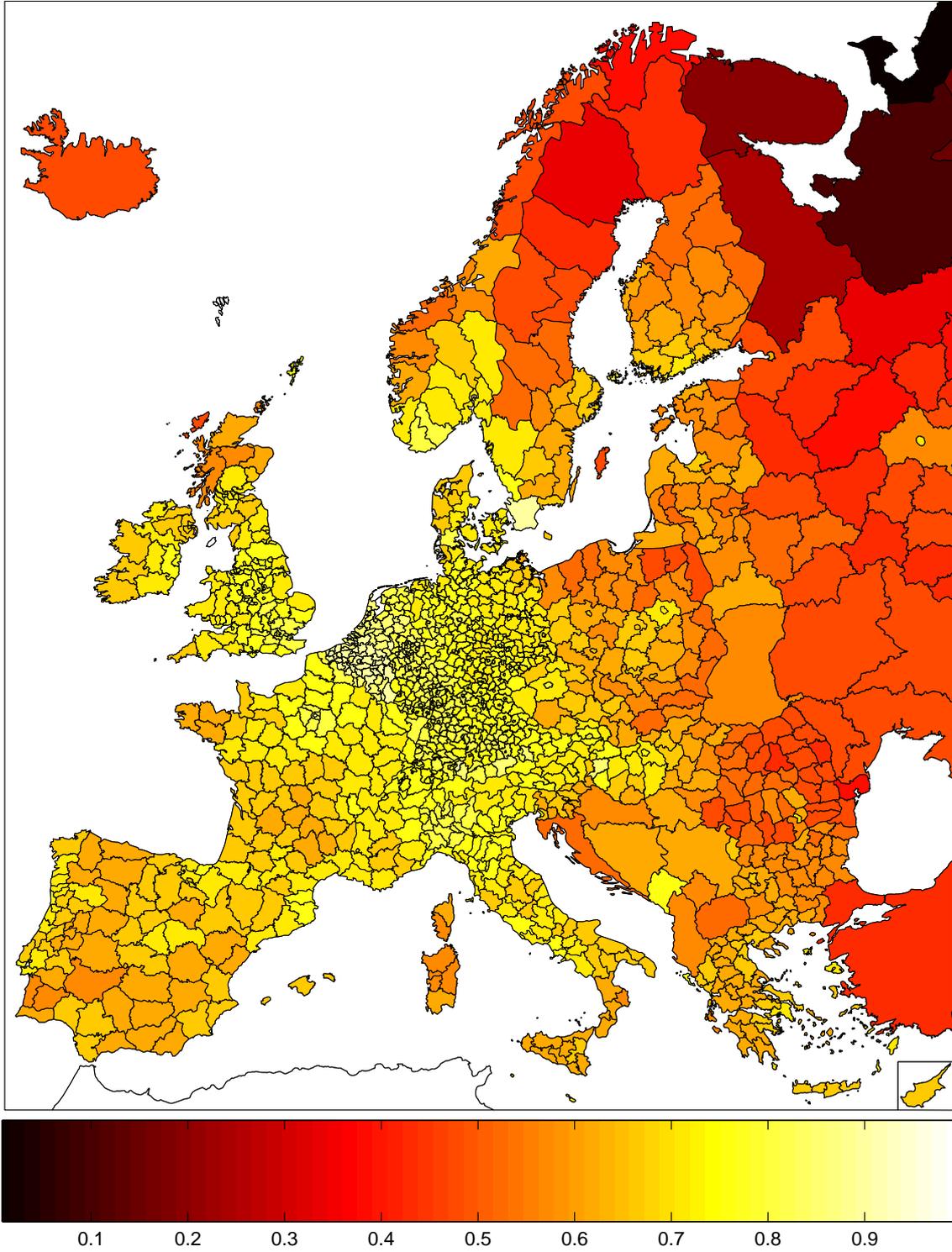


Figure 6.4: Accessibility Measures 2002 (Cross-Section Specification 1)



Implied values are $\sigma = 12$ and $\zeta = 0.01$. Data are transformed to the $[0,1]$ -scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

Figure 6.5: Accessibility Measures 2002 (Cross-Section Specification 2)



Implied values are $\sigma = 20.5$ and $\zeta = 0.01$. Data are transformed to the $[0,1]$ -scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

are rather assigned to capital regions than to the respective national peripheries. This fact can be easily seen in the case of Moscow, Bucharest, and Helsinki.

6.2.3 Sensitivity Analysis

Sensitivity analysis is carried out in two steps: Initially, we check the behavior of estimated parameters and coefficients over time by presenting estimates for the year 1995. This allows us to reveal whether parameter values tended to raise or to decline in the considered time period – even though the time interval is quite short. Falling values for σ would imply that consumers’ “love of variety” for manufactured varieties increased over time, i.e. that they become less substitutable. Because of the limited time horizon, results should be carefully interpreted.

In a second step, we analyze the functional sensitivity of the transport cost term. This is implied by our assumption of a value $\kappa = 0.5817$ (see chapter 4). It could be too restrictive to simply assume such a value since this parameter largely drives estimates of the nonlinearly estimated parameters, most of all the distance decay ζ . Alas, a different κ -value changes the shape of the transport cost and distance functions considerably. Because of this effect, we run sensitivity regressions using values of $\kappa = 0.3$ and 0.9 , respectively, in order to test for effects of a κ -shift in both directions.¹⁰

The results from our sensitivity analysis concerning the time dimension are reported in table 6.4. In this table, we concentrate on models which include productivity controls and only provide three models per specification. The results described before are confirmed and coefficient changes are quantitatively small. Estimates for σ seem to have risen over time whereas there is no clear evidence whether the distance decay rose or declined over time. Because the changes with respect to the nonlinear parameters are quite small and mixed with respect to signs, we cannot deduce a

¹⁰Another point to raise for sensitivity analysis could be that our set of control variables might not be complete, partly due to the fine regional disaggregation level chosen. On a higher geographic scale such as the NUTS2 level, we could gather more controls such as patent applications, female participation rates, age structure of the population, or education levels. However, these additional variables are only available from 1999 on. Consequently, we could only rely on a shorter time horizon, namely 1999-2002. Control regressions on the NUTS2 level (not reported here) provide us with two additional interesting insights. First, the only additional controls that exhibit a significant impact are female participation rates, high education levels, and patent applications, all of which have a significantly positive impact. However, they do not change the other conclusions derived so far. Second, in the NUTS2 setting, we had to exclude a large number of regions from the sample if we wanted to account for the additional controls, thus reducing the sample from 287 to 194 regions. As the degrees of freedom and the time horizon on NUTS2 level would shorten considerably and would not outweigh the gain of three more significant controls, we prefer to stay with our disaggregated data set and postpone the solution to the problem of omitted variable bias to our panel estimations.

clear cut statement concerning their movement over time. Thus, figures of the functional shape are not shown again. For specification 1, ζ rose for all specifications whereas it sharply falls for the specification 2-model in column (6). Surprisingly, in column (6) the LR-Test does reject a significant impact of accessibility on regional income levels. An explanation could be, that a prominent part of the agglomeration effects is already captured by the variable *urban*. In the spatial model, it turns out to be highly significant with a positive sign whereas it is insignificant in the models described in columns (4) and (5). For models of specification 1 it even had a significant but negative impact on regional income differences. After all, we regard the insignificant impact of market accessibility displayed in column (6) as an outlier.

Results for sensitivity analysis on the functional form of transport costs are displayed in table 6.5, where we report several variations of specification 1.¹¹ Column (1) shows results on varying the κ -value to 0.3 in the NLS approach including controls. Columns (2) and (3) report sensitivity analysis for the NLS-IV model and κ -values of 0.3 and 0.9. In analogy to this, we report spatial models in columns (4) to (6). With respect to the coefficients of the linear variables, we see that signs and coefficients magnitudes are robust to the sensitivity setups. All test statistics lead to identical conclusions as in the section above.

With respect to the nonlinearly estimated parameters, expectations are matched. Estimates for σ are stable, even though they decrease more in the IV case. As they should, the ζ -values rise when decreasing κ to 0.3, and decrease when κ is increased to 0.9. Surprisingly, the SGMM-IV approach is rejected by the significant test for overidentifying restrictions in column (6). The opposite is true for the model in columns (2), (3), and (5) where the instrumentation strategy turns out to be valid. Therefore, we trust in our instrumentation strategy.

We visualize the impact of changing κ -values for the NLS-IV model in figure 6.6. Here, the implied distance and transport cost functions are depicted. For all three κ -values we observe the assumed convex distance function (implying that the distance decay is decreasing with increasing distance), as well as the concave transport cost function (implying that transport costs rise less the greater the distance between two locations is). Curvatures for the sensitivity results are flatter than for the assumed baseline value $\kappa = 0.58$. This implies a lower distance decay and lower transport cost mark-ups. The respective outcomes for the spatial instrumentation models are shown in figure 6.7. For $\kappa = 0.3$ and $\kappa = 0.5817$, the functions behave similarly to those resulting from the NLS-IV models. However, for $\kappa = 0.9$ the distance decay becomes far steeper than before. Moreover, the transport cost function becomes convex. This

¹¹Results for specification 2 are not listed in detail. They also show the familiar patterns.

Table 6.4: Cross-Section Results: Variation over Time

Model	<i>NLS</i>	<i>NLSIV</i>	<i>SGMM</i>	<i>NLS</i>	<i>NLSIV</i>	<i>SGMM</i>
Proxy	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>	<i>gdph_{corr}</i>
Variable	$\omega_{r,1995}$	$\omega_{r,1995}$	$\omega_{r,1995}$	$\omega_{r,1995}$	$\omega_{r,1995}$	$\omega_{r,1995}$
<i>n</i>	1283	1283	1283	1283	1283	1283
σ	10.3962 (0.5297)	7.9993 (0.3142)	15.9601 (1.1052)	20.6989 (2.3765)	11.2792 (3.4681)	12.2824 (2.4109)
ζ	0.0205 (0.0018)	0.0117 (0.0018)	0.0142 (0.0012)	0.0128 (0.0014)	0.0052 (0.0010)	0.0007 (0.0006)
Constant	-1.7918*** (0.3585)	-2.4155*** (0.3486)	-0.5569*** (0.3710)			
East	-0.4097*** (0.0246)	-0.2625*** (0.0277)	-0.6859*** (0.0250)			
Capital	0.2712*** (0.0634)	0.4452*** (0.0660)	0.1779*** (0.0406)	0.2611*** (0.0412)	0.2849*** (0.0401)	0.2631*** (0.0365)
Sea	0.0597*** (0.0169)	0.0991*** (0.0174)	0.0269 (0.0185)	0.0047 (0.0145)	0.0162 (0.0150)	-0.0136 (0.0163)
Urban	-0.0795*** (0.0108)	-0.0668*** (0.0125)	-0.0219*** (0.0081)	0.0039 (0.0089)	0.0172 (0.0108)	0.0423*** (0.0079)
Kwh	-0.3746*** (0.0510)	-0.2729*** (0.0510)	-0.5390*** (0.0531)	-0.3288*** (0.0913)	-0.2841*** (0.0999)	-0.3737*** (0.1125)
Ind	0.0437* (0.0254)	0.0459* (0.0271)	0.0573** (0.0212)	0.0581*** (0.0223)	0.0441* (0.0232)	0.0775*** (0.0188)
Agrar	-0.0640*** (0.0094)	-0.0476*** (0.0123)	-0.0485*** (0.0055)	-0.0323*** (0.0066)	-0.0367*** (0.0071)	-0.0278*** (0.0050)
ϱ			0.6124***			0.4061***
Country	no	no	no	yes	yes	yes
ξ_{LR}	479.4***	272.3***	59.5***	133.1***	21.8***	2.4638
ξ_{MoranI}	0.4015***	0.5442***		0.1953***	0.2668***	
ξ_{SERR}	417.6***	957.2***		719.3***	268.1***	
ξ_{SLAG}	6.8383***	1016.1***		0.8519	226.1***	
ξ_{DWH}		-0.4804***			-0.4959***	
ξ_{OR} prob		0.7352			0.9987	
adj. R^2	0.7339	0.6918	0.8252	0.8701	0.8583	0.8751

White's heteroscedastic standard errors in parenthesis except for spatial models, where White's correction is not applied. Standard errors for instrumentation models are based on the adjusted instrumentation error. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, and * denotes significance at the 0.1 level. Tests for weak instruments are not shown in detail but clearly prove the power of the applied instruments. For ξ_{OR} only the t-probabilities are displayed. Spatial tests and models are performed using a spatial weights matrix with weights decreasing linearly from 1 to 0 while the centroid distance increases from 0 to 100 minutes travel time. Country dummy values are not reported in detail. The use of a full set of country dummies makes the use of an intercept term as well as *East* unnecessary.

Table 6.5: Cross-Section Results: Functional Sensitivity (Specification 1)

Model	<i>NLS</i>	<i>NLSIV</i>	<i>NLSIV</i>	<i>SGMM</i>	<i>SGMMIV</i>	<i>SGMMIV</i>
κ	0.3	0.3	0.9	0.3	0.3	0.9
Variable	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$	$\omega_{r,2002}$
n	1283	1283	1283	1283	1283	1283
σ	11.3345 (0.5688)	9.4614 (0.4557)	10.1738 (0.5896)	15.9637 (1.1461)	9.0992 (0.6004)	9.8535 (0.5355)
ζ	0.0960 (0.0097)	0.0238 (0.0074)	0.0003 (0.0001)	0.1077 (0.0099)	0.0292 (0.0080)	0.0040 (0.0001)
Constant	-2.7573*** (0.3098)	-2.6738*** (0.2979)	-2.4049*** (0.2974)	-1.7558*** (0.3476)	3.1294*** (0.4496)	7.3091*** (0.6825)
East	-0.4267*** (0.0189)	-0.4356*** (0.0210)	-0.4538*** (0.0210)	-0.6315*** (0.0228)	-1.1776*** (0.0304)	-1.5469*** (0.0470)
Capital	0.3497*** (0.0506)	0.5130*** (0.0516)	0.4979*** (0.0515)	0.2398*** (0.0382)	0.2741*** (0.0450)	0.1086 (0.0691)
Sea	0.0723*** (0.0162)	0.0483*** (0.0165)	0.0557*** (0.0164)	0.0426** (0.0173)	0.0713*** (0.0217)	0.0294 (0.0333)
Urban	-0.0750*** (0.0087)	-0.0492*** (0.0090)	-0.0516*** (0.0090)	-0.0388*** (0.0079)	-0.0386*** (0.0100)	-0.0433*** (0.0146)
Kwh	-0.1888*** (0.0437)	-0.1913*** (0.0423)	-0.2344*** (0.0423)	-0.3202*** (0.0495)	-1.0211*** (0.0639)	-1.7629*** (0.0976)
Ind	0.0509** (0.0214)	0.0997** (0.0227)	0.0812*** (0.0227)	0.0534 (0.0181)	0.0525** (0.0219)	0.0302 (0.0371)
Agrar	-0.0807*** (0.0081)	-0.0729*** (0.0089)	-0.0739*** (0.0089)	-0.0601*** (0.0056)	-0.0728*** (0.0068)	-0.0552*** (0.0096)
ϱ				0.5960***	0.6965***	0.7029***
ξ_{LR}	391.7***	219.1***	219.2***	40.0***	70.4***	109.6***
ξ_{MoranI}	0.3516***	0.4245***	0.4228***			
ξ_{SERR}	647.9***	678.5***	672.9***			
ξ_{SLAG}	1.4550	696.2***	683.0***			
ξ_{DWH}		-0.3433***	-0.3130***		-15.56***	-0.7703***
ξ_{OR} prob		0.9647	0.9781		0.4467	0.0000
adj. R^2	0.7351	0.7155	0.7162	0.8113	0.7935	0.6675

White's heteroscedastic standard errors in parenthesis except for spatial models, where White's correction is not applied. Standard errors for instrumentation models are based on the adjusted instrumentation error. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, and * denotes significance at the 0.1 level. Tests for weak instruments are not shown in detail but clearly prove the power of the applied instruments. For ξ_{OR} only the t-probabilities are displayed. Spatial tests and models are performed using a spatial weights matrix with weights decreasing linearly from 1 to 0 while the centroid distance increases from 0 to 100 minutes travel time. Country dummy values are not reported in detail. The use of a full set of country dummies makes the use of an intercept term as well as *East* unnecessary.

Figure 6.6: Functional Sensitivity of Distance & Transport Cost Functions (Specification 1)

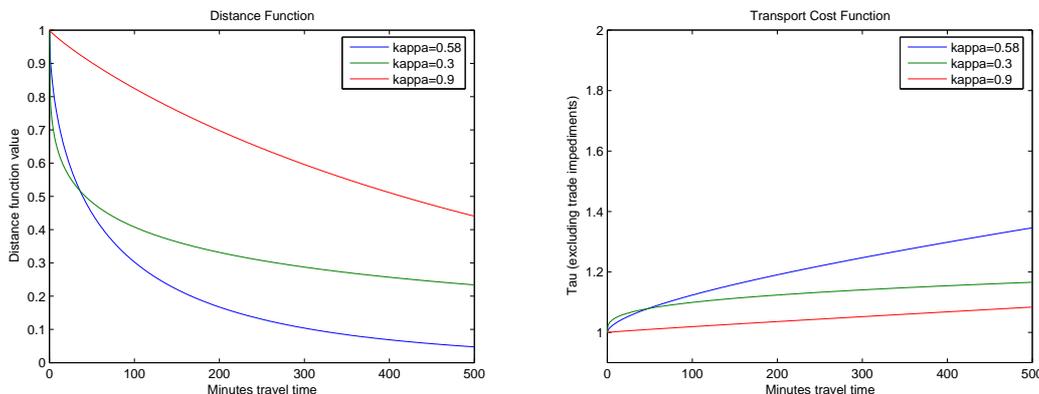
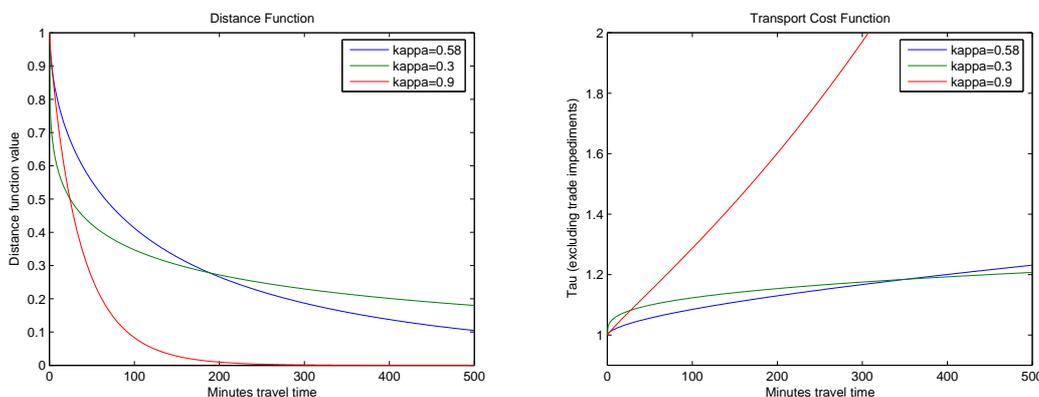


Figure 6.7: Functional Sensitivity of Distance & Transport Cost Functions (Spatial Model Specification 1)



implies an implausibly big transport-cost markup of roughly 100% within less than 300 minutes travel time. This enormous change is due to the estimate for ζ which appears to be ten times higher than for the NLS-IV models.

6.3 Panel Results

In this section, we present the outcomes of our panel estimations. On the one hand, we present the results of the random effects models, including our panel instrumentation approach. On the other hand, we briefly comment on our attempts to estimate fixed effects and SGMM models. As pointed out below, results from such an analysis exhibit several weaknesses.

The random effects results are based on 10264 observations and are summarized in table 6.6.¹² Macroeconomic shocks are absorbed by time dummies. Columns (1) to

¹²In the IV case, we can rely on 7698 observations.

(3) comprise the simple NLS and the NLS-IV estimates for specification 1, columns (4) to (6) enclose the respective results for specification 2. For all columns, spatial tests are not reported in detail but turn out to be highly significant.¹³

The LR-Test reveals that the impact of accessibility remains significant but decreases considerably in absolute terms compared to the cross-section estimates. Contrary to the latter, the insignificant test for overidentifying restrictions rejects the validity of our instrumentation approach for specification 1. However, the test statistic proves the suitability of our instrumentation strategy for specification 2. Having a look at first stage regression results (not reported in detail), we confirm a significant impact of the instrument on the endogenous variable for both specifications. But due to the ambiguous test statistic, we are sceptic if our instrument is really applicable. Unfortunately, we fail to provide an alternative instrument for our panel setup.

Compared to the aforementioned results, the coefficients of the linear model parameters stay stable with respect to their signs except for the model in column (6). Here, signs for *Sea*, *Urban*, and *Kwh* are in contrast to our expectations. In total, all linear coefficients become closer to zero. The reason is simple: since we control for region specifics, we capture the major influences on the regional income level by random individual intercepts. The suppressed individual coefficients are all highly significant and relatively high by value compared to the other coefficients. Country effects have a standard deviation ranging from 0.3 (mean = -0.7) for the logarithmic model in column (6) to 0.43 (mean = -0.48) for the model in column (2). These standard deviations imply considerably large productivity differentials. A difference of 0.3 between country effects corresponds to a productivity ratio of $\exp(0.3) = 1.35$, ceteris paribus. For a difference of 0.55 the ratio is 1.54.

The model fit increases to roughly 0.9 for all reported models. In the panel setup, σ -values are similar to the cross-sectional analysis and now are fluctuating between 8 and 13. Estimates for ζ are remarkably smaller than for the former strategy. The implied transport cost and distance functions are shown in figure 6.8. They reveal the familiar pattern but appear to be flatter than for the aforementioned estimation results.

In figures 6.9 and 6.10, the regional accessibility measures implied by the random effects panel approaches towards both specifications are visualized. In order to compare market proximity figures to those described before, we present the accessibility measures for 2002. The inclusion of region specific effects does not change the landscape of regional demand and supply prices in principle. The center-periphery

¹³For the extension of spatial LM-Error-Tests to the panel case, see Baltagi *et al.* (2003).

Table 6.6: Panel Results: Random Effects Models

Model	NLS_{RE}	NLS_{REIV}	NLS_{REIV}	NLS_{RE}	NLS_{REIV}	NLS_{REIV}
Proxy	$gdph_{corr}$	$gdph_{corr}$	$gdph_{corr}$	$gdph_{corr}$	$gdph_{corr}$	$gdph_{corr}$
Variable	$\omega_{r,95-02}$	$\omega_{r,97-02}$	$\omega_{r,97-02}$	$\omega_{r,95-02}$	$\omega_{r,97-02}$	$\omega_{r,97-02}$
κ	0.58	0.58	0.58	0.58	0.58	0.58
n	10264	7698	7698	10264	7698	7698
σ	12.9545 (0.6011)	8.7752 (0.4201)	8.6060 (0.4317)	11.9397 (0.7464)	10.3339 (0.8999)	8.3560 (0.5258)
ζ	0.0035 (0.0007)	0.0083 (0.0001)	0.0073 (0.0001)	0.0015 (0.0008)	0.0116 (0.0001)	0.0053 (0.0001)
Constant	-3.0589*** (0.0318)	-3.5210*** (0.0259)	-2.1736*** (0.3763)	-2.6841*** (0.0974)	-3.0399*** (0.0719)	-3.3888*** (0.7289)
East	-0.0872*** (0.0037)	-0.0993*** (0.0047)	-0.0865*** (0.0065)			
Capital			0.1146*** (0.0192)			0.0728*** (0.0177)
Sea			0.0249*** (0.0037)			-0.0064 (0.0043)
Urban			-0.0137*** (0.0023)			0.0009 (0.0027)
Kwh			-0.0453*** (0.0107)			0.0040 (0.0220)
Ind			0.0900*** (0.0213)			0.0797*** (0.0233)
Agrar			-0.0803*** (0.0107)			-0.0729*** (0.0107)
Time	yes	yes	yes	yes	yes	yes
Region	yes	yes	yes	yes	yes	yes
Country	no	no	no	yes	yes	yes
ξ_{LR}	324.4***	107.5***	144.7***	274.3***	92.8***	163.5***
ξ_{DWH}		-0.0993***	-0.0869*		-0.1934***	-0.0744
ξ_{OR} prob		0.0068	0.0000		0.4327	0.7785
adj. R^2	0.9204	0.9051	0.9075	0.9239	0.9071	0.9124

White's heteroscedastic standard errors in parenthesis. Standard errors for instrumentation models are based on the adjusted instrumentation error. *** denotes significance at the 0.01 level, ** denotes significance at the 0.05 level, and * denotes significance at the 0.1 level. Tests for weak instruments are not shown in detail but clearly prove the power of the applied instruments. For ξ_{OR} only the t-probabilities are displayed. Random region effects as well as coefficients for time and country dummies not reported in detail. Due to the intercept term, country dummies are included as a set for all countries but one and have to be interpreted as adjustment factors for the intercept term. Spatial tests are not reported in detail but turn out to be highly significant.

Figure 6.8: Distance & Transport Cost Functions (Panel Models)

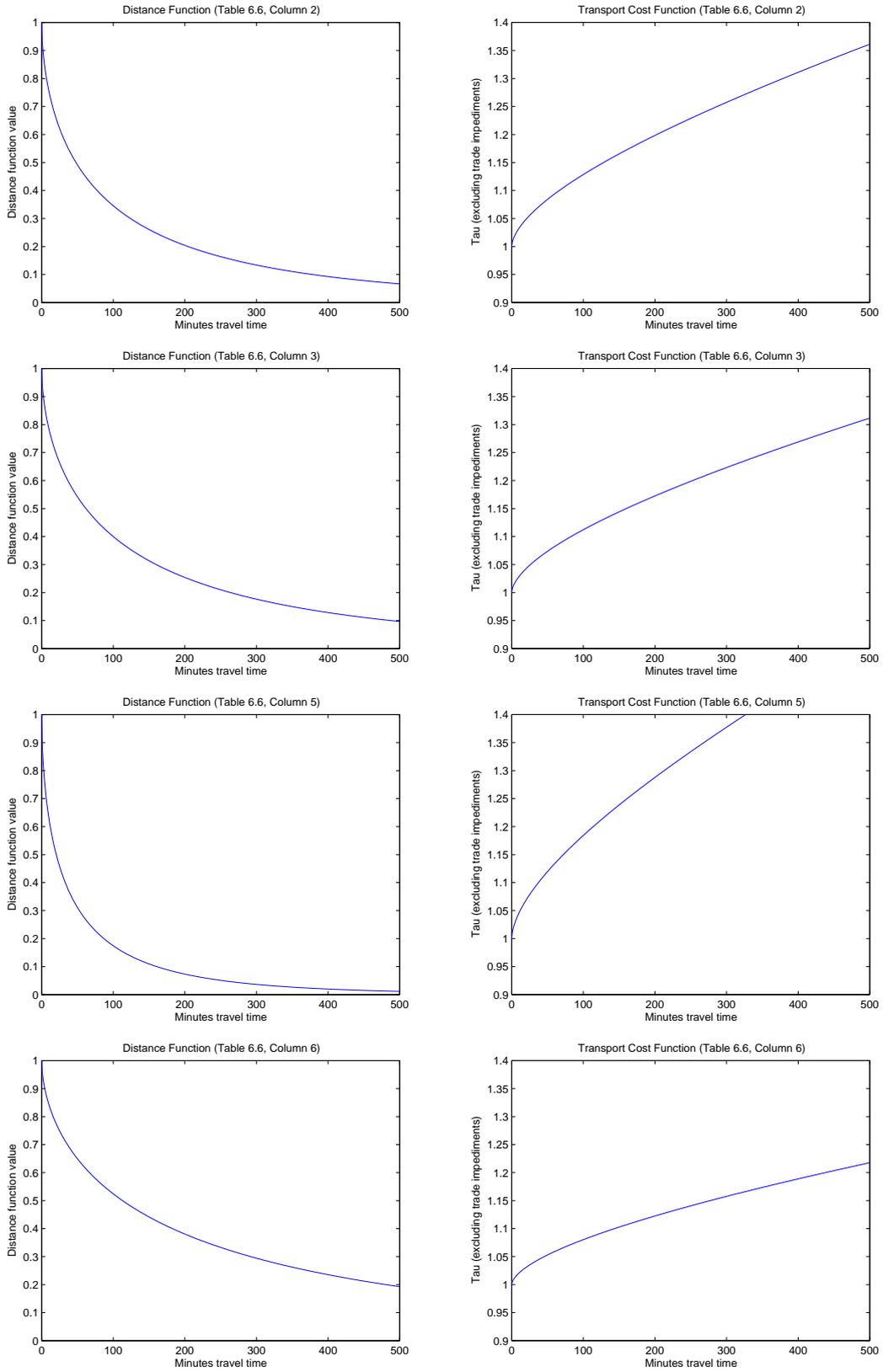
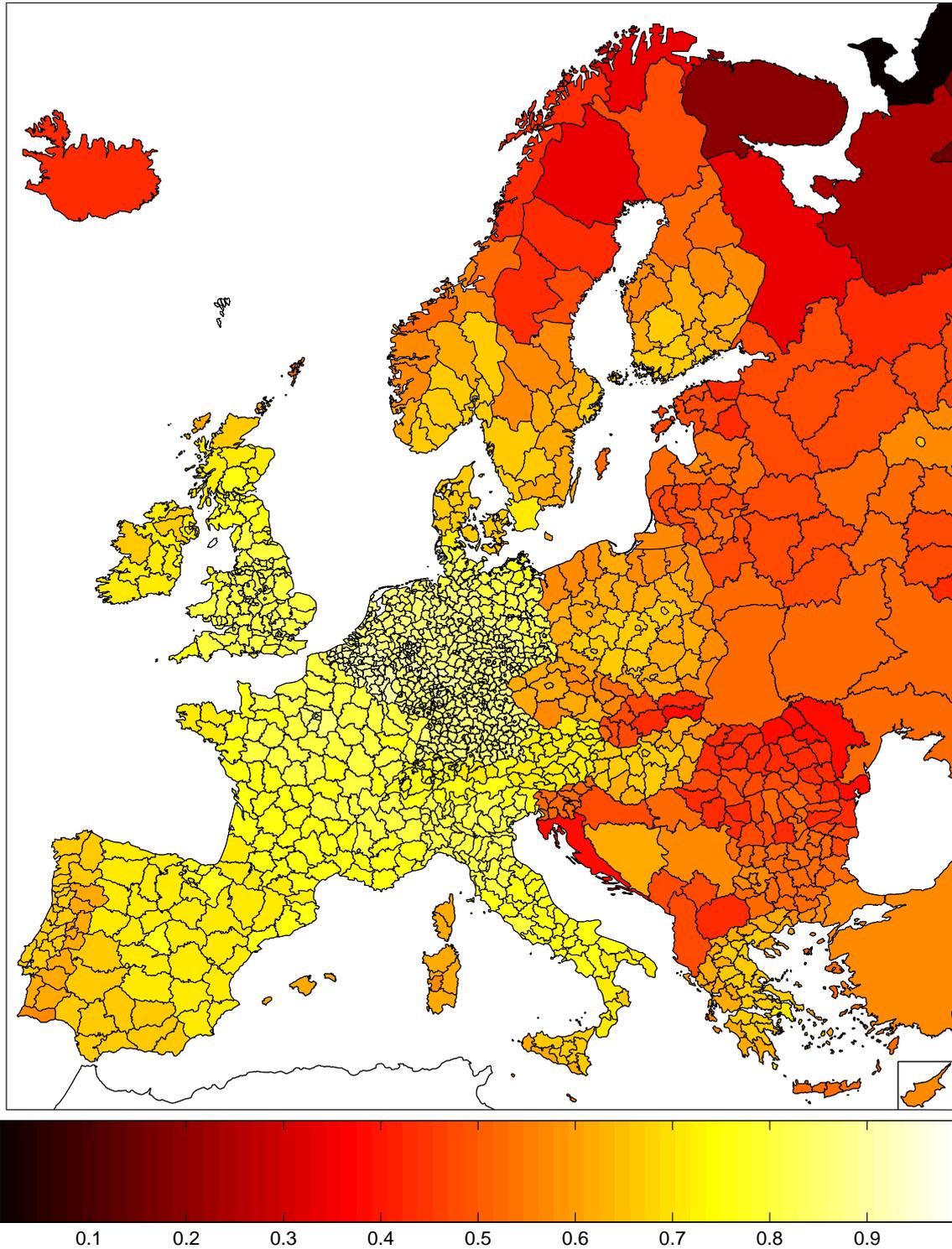
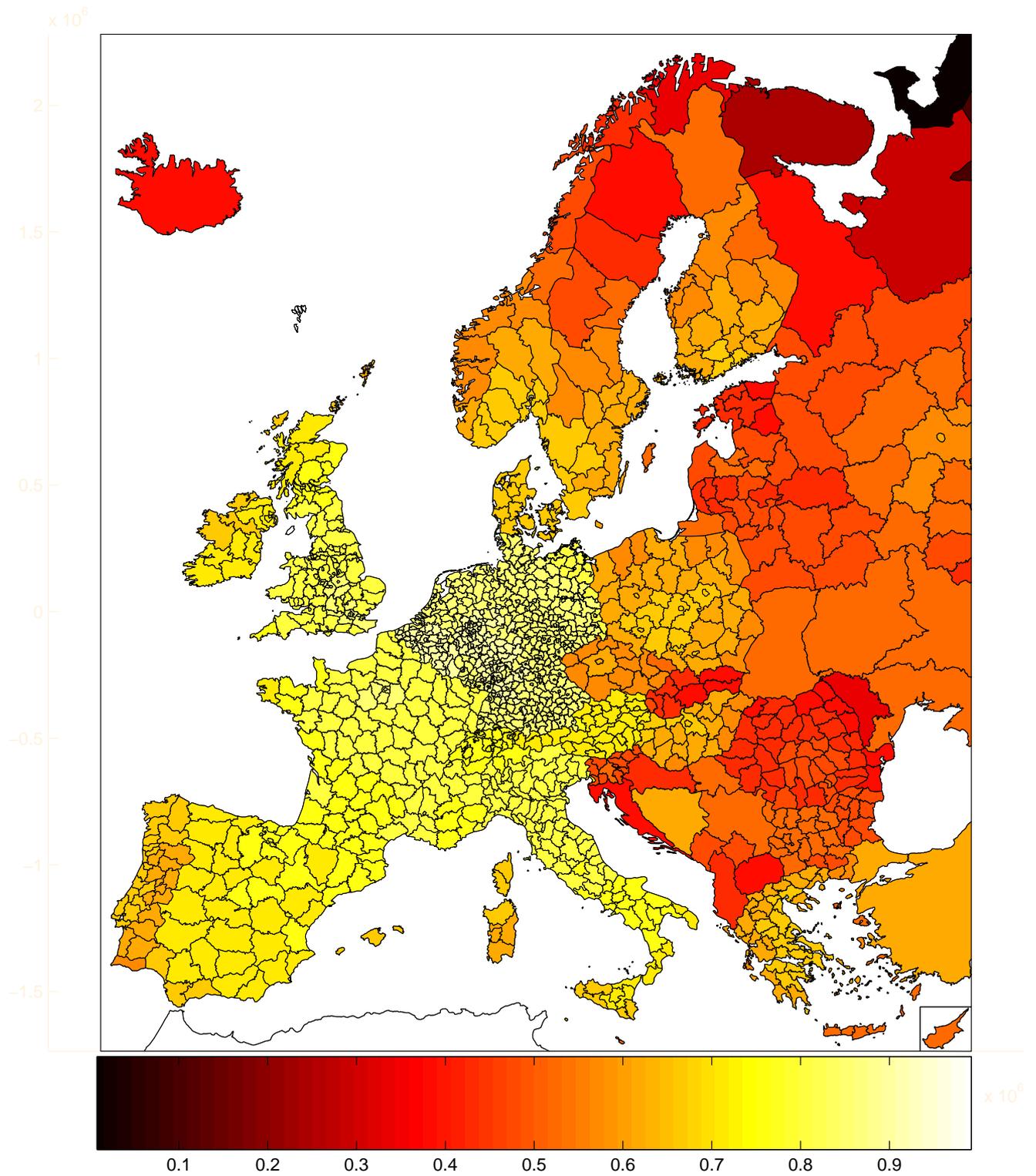


Figure 6.9: Accessibility Measures 2002 (Panel Specification 1)



Results are from Panel specification 1, using the NLS-IV estimator from column (3) in table 6.6. Implied values $\sigma = 8.6$, and $\zeta = 0.007$. Data are transformed to the $[0,1]$ -scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

Figure 6.10: Accessibility Measures 2002 (Panel Specification 2)



Results are from Panel specification 2 using the NLS-IV estimator from column (6) in table 6.6. Implied values $\sigma = 8.3$, and $\zeta = 0.005$. Data are transformed to the $[0,1]$ -scale by segmenting the data into 60 intervals of identical length, with 0 being the minimum value and 1 the maximum value of the original data.

pattern is observable on both the national and international level. But we are observing sharp cross-border differences in regional accessibility now. For instance, Germany in total is assigned to have a very good relative accessibility. Hence, for smaller distance parameter estimates we clearly reveal that our accessibility measure rather explains national than regional income differences. On the contrary, a variation of σ does not result in significantly different landscapes. Note, that the estimates for ζ underlying figures 6.9 and 6.10 are comparable to the outcomes of the NLS-IV model from our cross-sectional analysis.

Results of sensitivity analysis with regard to the form of the transport cost function and results of spatial models are not shown in detail. When applying spatial models, we were not able to recover estimates of $\zeta > 0$. Recall that the latter is theoretically required. At the same time, estimates for σ are far higher than for the presented models. This means that estimating complicated models which correct for the apparent spatial autocorrelation is not suitable for our empirical panel strategy. Sensitivity analysis with respect to κ failed to provide estimates of $\zeta > 0$ for specification 2, too. But estimates for σ stay in the ranges reported in table 6.6. However, sensitivity analysis confirms the results of the cross-section case for specification 1. Because coefficients are stable compared to the already presented results, we do not show them in detail again.

Concerning fixed effects panel estimates, we only give a brief summary. In principle, this approach comes at cost of leveling out all time-invariant explanatory variables. Logically consistent, it is only leaving us with industry and agriculture shares of GVA besides the region-specific intercepts. Hence, as both possible time-invariant dummy measures of the technology level cannot be included, no distinction between specifications 1 and 2 has to be made.¹⁴ The argument in favor of this approach is that we test for changing accessibility in a constant geographic structure. We do not show results in detail because our empirical strategy does not yield estimates of $\zeta > 0$.¹⁵ Estimates for σ are a bit lower compared to the results of the random effects models and range between 6 and 8.

6.4 Chapter Summary

In this chapter, we showed that regional income differences across Europe can be explained by regional accessibility to a limited extent. We found evidence for at least

¹⁴In this setting, we do not estimate spatial models, since the method of fixed effects does only provide us with starting values for σ_v^2 and not for σ_ϵ^2 .

¹⁵The same holds true when estimating cross-section models by first-differences.

some positive impact across almost all cross-section-models estimated. However, the test statistic decreases when instrumenting and especially when correcting for spatial misspecification. Here, the LR test is quite small and in rare cases the impact of regional supply and demand prices turns out to be even insignificant. Moreover, market proximity is not the only important factor behind the income differences. Especially the coefficients accounting for the industry structure of a region and the dummy for capital regions were highly significant across all estimates. Unobservable country effects appeared to be highly significant, too.

Applying panel data techniques confirms these outcomes. Here, the coefficients for control variables are close to zero. Hence, panel data setups reveal that a large part of the spatial income differences is explained by unobserved region characteristics instead of relative regional prices or other explaining variables. However, we fail to provide valid results for spatial econometrics models in the panel setup because parameter estimates for ζ are inconsistent with theory. Additionally, our instrumentation strategy performs weakly for some of the estimated panel models. We were not able to find alternative instruments which mitigate this shortcoming. Therefore, we need to be careful when interpreting our panel results because we could not correct for the spatial autocorrelation in the data.

Across most of the presented results, it appeared that σ is clustered within a range of 8 to 15. This is at the upper end of the range implied by the studies reviewed in chapter 2 and suggests that tradable goods are better substitutes than expected. Following our cross-section analysis, ζ is close to 0.01 on average, implying a trade cost of about 45% of the product value for 500 minutes transport distance, which is a sensible magnitude. In the majority of panel estimates the estimated trade cost drops to a range between about 15% and 50% within 500 minutes transport distance. Altering the value for κ only influences the results for ζ , as it should, but not for σ . In general, it seems that income differences between countries are better explained by the model than within countries, because including country dummies sharply reduces the explanatory power of accessibility.

Chapter 7

Conclusions

The thesis' task was to test whether models of *New Trade Theory* are valuable tools for explaining spatial income differences across European regions. The key hypothesis of these models is that a region's relative location is a major explanatory factor of its economic strength. We tested these models, using an integrated spatial CGE and estimation approach, which calibrates the underlying general equilibrium model and estimates a wage equation simultaneously.

This strategy eliminates a couple of problems that occur in state-of-the-art approaches: First, we depend less on data restrictions since regional trade and prices as well as bilateral trade impediments are the outcome of our model calibration. Second, we unify the approaches of Hanson (2005) and Redding & Venables (2004). That means, we do not only test single key relationships of NEG models but account for all dimensions of the model. Third, we do not need to introduce restrictive assumptions about real wage equalization across regions. Moreover, we apply a well defined transport cost function, which consists of a distance-related component as well as one accounting for bilateral trade impediments. We apply our approach to the regional context of the EU-27 and use disaggregated data at the same time. Compared to analyses with a restricted regional focus, our approach derives more general results which are not influenced by, for example, a national or Western European perspective. Additionally, our econometric strategy is comprehensive. It allows us to rely on both supply and demand prices when measuring market proximity. We calibrate the model for several time periods in our panel specifications. In fact, this strategy accounts for changes in the equilibrium solution over time. Moreover, several econometric problems are eliminated by usage of instrumentation, spatial-econometrics and panel methods.

The major results of our study are the following: The influence of regional accessibility as one major explanatory variable of regional income differences is proven.

From a cross-sectional perspective, the impact of market proximity is highly significant. This result is robust to the application of instrumentation models. Using spatial econometrics models yields quite low values for the respective test statistic but the influence of regional accessibility remains significant for most models. Including alternative explaining forces into the analysis does not change these outcomes. Estimates from our random effects approach confirm the cross-sectional evidence.

These conclusions are also robust to the variation of the transport cost function. An alteration of parameter κ which shapes the transport cost function naturally changes the magnitude of estimates for the distance parameter ζ . However, estimates for the elasticity of substitution σ turn out to be robust to this sensitivity analysis. Moreover, the alteration of κ does neither change the conclusions with regard to the impact of accessibility on the income level which remains highly significant.

We estimate σ -values which are at the upper end of the range implied by existing studies. For most of our specifications σ fluctuates between 8 and 15. This indicates that tradable goods are better substitutes and that trade margins for Europe are not as high as suggested in the survey by Anderson & van Wincoop (2004). Our estimates imply transport cost markups between 15% and 30% within 400 minutes travel time.

In the introductory chapter, we raised the question whether regional policy should or should not follow the key predictions of spatial trade models. We showed that a region's market proximity plays an important role in the explanation of regional income levels. Hence, improving transport infrastructure and further integrating the *European Single Market* should help to enhance a region's wealth at least to some extent. At the same time, we identified other important adjustment factors. For instance, politicians should further stress the importance of indicators describing the industry structure. We showed, that high shares of industry production and low shares for the agricultural sector go along with higher incomes. Therefore, concentrating political efforts on the amelioration of regional accessibility would be myopic if one aims at reducing regional income differences.

Bibliography

- M. AMITI (1999). Specialization Patterns in Europe. *Weltwirtschaftliches Archiv* **135**, 573–593.
- J.E. ANDERSON & E. VAN WINCOOP (2003). Gravity with Gravititas: A Solution to the Border Puzzle. *American Economic Review* **93**, 170–192.
- J.E. ANDERSON & E. VAN WINCOOP (2004). Trade Costs. *Journal of Economic Literature* **72**, 691–751.
- L. ANSELIN (1988a). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis* **20**, 1–17.
- L. ANSELIN (1988b). *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.
- L. ANSELIN & R.J.G.M. FLORAX (1995). Small Sample Properties of Tests for Spatial Dependence in Regression Results: Some Further Results. In *New Directions in Spatial Econometrics*, L. ANSELIN & R.J.G.M. FLORAX, editors, 21–73. Springer, Berlin.
- L. ANSELIN & S. REY (1991). Properties of Tests for Spatial Dependence in Linear Regression Models. *Geographical Analysis* **23**, 112–131.
- M. ARRELANO & S. BOND (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58**, 277–297.
- D.B. AUDRETSCH & M.P. FELDMAN (1996). R&D Spillovers and the Geography of Innovation and Production. *American Economic Review* **86**, 630–640.
- S.L. BAIER & J.H. BERGSTRAND (2001). The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity. *Journal of International Economics* **53**, 1–27.

- R.E. BALDWIN, R. FORSLID, P. MARTIN, G.I.P. OTTAVIANO & F. ROBERT-NICOUD (2003). *Economic Geography and Public Policy*. Princeton University Press, Princeton.
- R.E. BALDWIN, J.F. FRANCOIS & R. PORTES (1997). The Costs and Benefits of Eastern Enlargement: The Impact on the EU and Central Europe. *Economic Policy* **12**, 125–176.
- B.H. BALTAGI, S.H. SONG & W. KOH (2003). Testing Panel Data Regression Models with Spatial Error Correlation. *Journal of Econometrics* **117**, 123–150.
- S. BRAKMAN, H. GARRETSEN & C. VAN MARREWIJK (2001). *An Introduction to Geographical Economics*. Cambridge University Press, Cambridge.
- S. BRAKMAN, H. GARRETSEN & M. SCHRAMM (2004). The Spatial Distribution of Wages: Estimating the Helpman-Hanson Model for Germany. *Journal of Regional Science* **44**, 437–466.
- S. BRAKMAN, H. GARRETSEN & M. SCHRAMM (2006). Putting New Economic Geography to the Test: Free-ness of Trade and Agglomeration in the EU Regions. *Regional Science and Urban Economics* **36**, 613–635.
- J. BRÖCKER (1984). How Do International Trade Barriers Affect Interregional Trade? In *Regional and Industrial Development Theories*, A.E. ANDERSSON, W. ISARD & T. PUU, editors, 219–239. Elsevier, Amsterdam.
- J. BRÖCKER (1995). Chamberlinian Spatial Computable General Equilibrium Modelling: A Theoretical Framework. *Economic Systems Research* **7**, 137–149.
- J. BRÖCKER (1998a). *Spatial Effects of Transeuropean Networks: Preliminary Results from a Spatial Computable General Equilibrium Model*. Diskussionsbeiträge aus dem Institut für Wirtschaft und Verkehr 4.98, Technische Universität Dresden.
- J. BRÖCKER (1998b). How Would an EU- Membership of the Visegrád-Countries affect Europe's Economic Geography? *Annals of Regional Science* **32**, 91–114. Erratum (2000), 34, 469–471.
- J. BRÖCKER (2003). Regionale Wirkungen der Europäischen Währungsunion. *Jahrbuch für Regionalwissenschaft* **23**, 3–22.
- J. BRÖCKER (2005). *Spatial Effects of Transport Initiatives: An Update*. Paper presented at the Kiel Summer Workshop on Trade and Location, June 3rd-5th, Kiel.

- J. BRÖCKER (2006). *Parametric and Non-Parametric Accessibility Measurement*. Paper presented to the 46th ERSAs Congress, Volos, Greece.
- J. BRÖCKER, U. KRETSCHMER, C. SCHÜRMAN, K. SPIEKERMANN, D. STELDER & M. WEGENER (2002). *The IASON Common Spatial Database*. IASON Deliverable 3, Christian-Albrechts-Universität zu Kiel/Institut für Raumplanung, Kiel/Dortmund.
- J. BRÖCKER & H.C. ROHWEDER (1990). Barriers to International Trade: Methods of Measurement and Empirical Evidence. *Annals of Regional Science* **24**, 289–305.
- H. BREINLICH (2006). The Spatial Income Structure in the European Union - What Role for Economic Geography? *Journal of Economic Geography* **6**, 593–617.
- M. BRÜLHART, M. CROZET & P. KÖNIG (2004). *Enlargement and the EU Periphery: The Impact of Changing Market Potential*. HWWA Discussion Paper 270, Hamburg.
- M. BRÜLHART & P. KÖNIG (2006). New Economic Geography Meets COMECON: Regional Wages and Industry Location in Central Europe. *Economics of Transition* **14**, 245–267.
- P. BURRIDGE (1980). On the Cliff-Ord Test for Spatial Autocorrelation. *Journal of the Royal Statistical Society B* **42**, 107–108.
- E. CHAMBERLIN (1933). *The Theory of Monopolistic Competition*. Harvard University Press, Harvard.
- P. CHESHIRE & D.G. HAY (1989). *Urban Problems in Western Europe: An Economic Analysis*. Unwyn Hyman, London.
- A. CICCONE & R.E. HALL (1996). Productivity and the Density of Economic Activity. *American Economic Review* **86**, 54–70.
- A. CLIFF & J.K. ORD (1972). Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis* **4**, 267–284.
- D.T. COE & E. HELPMAN (1995). International R&D Spillovers. *European Economic Review* **39**, 859–887.
- P.P. COMBES & M. LAFOURCADE (2001). *Transport Cost Decline and Regional Inequalities: Evidence from France*. CEPR Discussion Paper 2894, London.

- M. CROZET (2004). Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model. *Journal of Economic Geography* **4**, 439–458.
- J.P. DAMIJAN & C. KOSTEVC (2003). The Impact of European Integration on the Adjustment Pattern of Regional Wages in Accession Countries. In *The Emerging Economic Geography in EU Accession Countries*, I. TRAISTARU, P. NIJKAMP & L. RESMINI, editors, 372–404. Ashgate, Aldershot.
- R. DAVIDSON & J.G. MACKINNON (2004). *Econometric Theory and Methods*. Oxford University Press, New York.
- D.R. DAVIS & D.E. WEINSTEIN (1999). Economic Geography and Regional Production Structure: An Empirical Investigation. *European Economic Review* **43**, 379–407.
- D.R. DAVIS & D.E. WEINSTEIN (2003). Market Access, Economic Geography and Comparative Advantage: An Empirical Test. *Journal of International Economics* **59**, 1–23.
- B. N. DELAUNAY (1934). Sur la sphère vide. *Bulletin of Academy of Sciences of the USSR* **7**, 793–800.
- B.V. DIMARANAN (2006). *Global Trade, Assistance, and Production: The GTAP 6 Data Base*. Center for Global Trade Analysis, Purdue University, West Lafayette (IN).
- A.C. DISDIER & K. HEAD (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *Review of Economics and Statistics* **90**, 37–48.
- A.K. DIXIT & J. STIGLITZ (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review* **67**, 297–308.
- G. ELLISON & E.L. GLAESER (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* **105**, 889–927.
- G. ELLISON & E.L. GLAESER (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review Papers and Proceedings* **89**, 311–316.
- EUROPEAN COMMISSION (2004). *Third Report on Economic and Social Cohesion: Unity, Solidarity, and Diversity for Europe*. Brussels.

- EUROSTAT (1999). *Nomenclature of Territorial Units for Statistics (NUTS)*. Luxembourg.
- R.C. FEENSTRA (1994). New Product Varieties and the Measurement of International Prices. *American Economic Review* **84**, 157–177.
- R.C. FEENSTRA (2002). Border Effects and the Gravity Equation: Consistent Methods for Estimation. *Scottish Journal of Political Economy* **49**, 491–506.
- M.P. FELDMAN (1999). The New Economics of Innovation, Spillovers, and Agglomeration: A Review of Empirical Studies. *Economics of Innovation and New Technology* **8**, 5–25.
- R.J.G.M. FLORAX & S. REY (1995). The Impacts of Misspecified Spatial Interaction in Linear Regression Models. In *New Directions in Spatial Econometrics*, L. ANSELIN & R.J.G.M. FLORAX, editors, 111–135. Springer, Berlin.
- M. FUJITA, P. KRUGMAN & A.J. VENABLES (2001). *The Spatial Economy: Cities, Regions, and International Trade (First Paperback Edition)*. MIT Press, Cambridge (MA).
- J.L. GALLUP, J.D. SACHS & A.D. MELLINGER (1999). Geography and Economic Development. *International Regional Science Review* **22**, 179–232.
- R.E. HALL & C.I. JONES (1999). Why Do Some Countries Produce so Much More Output Than Others? *Quarterly Journal of Economics* **114**, 83–116.
- G. HANSON (2001). Scale Economies and the Geographic Concentration of Industry. *Journal of Economic Geography* **1**, 255–276.
- G. HANSON (2005). Market Potential, Increasing Returns, and Geographic Concentration. *Journal of International Economics* **67**, 1–24.
- C.D. HARRIS (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers* **44**, 315–348.
- J.A. HAUSMAN (1978). Specification Tests in Econometrics. *Econometrica* **46**, 1251–1271.
- J. HAVEMAN & D. HUMMELS (1999). Alternative Hypothesis and the Volume of Trade: Evidence on the Extend of Specialization. Mimeo, Sogang University Seoul. Internet: <http://www.sogang.ac.kr/~eysong/paper/Haveman.pdf>.

- K. HEAD & T. MAYER (2000). Non-Europe: The Magnitude and Causes of Market Fragmentation in Europe. *Weltwirtschaftliches Archiv* **136**, 285–314.
- K. HEAD & T. MAYER (2002a). Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade. CEPII Working Paper 2002-01, Paris.
- K. HEAD & T. MAYER (2002b). On the Pervasiveness of the Home Market Effect. *Economica* **69**, 371–390.
- K. HEAD & T. MAYER (2004). The Empirics of Agglomeration and Trade. In *Handbook of Regional and Urban Economics*, J.V. HENDERSON & J.F. THISSE, editors, volume 4, 2609–2669. North Holland, Amsterdam.
- K. HEAD & T. MAYER (2006). Regional Wage and Employment Responses to Market Potential in the EU. *Regional Science and Urban Economics* **36**, 573–594.
- K. HEAD & J. RIES (2001). Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade. *American Economic Review* **91**, 858–876.
- E.F. HECKSCHER (1919). The Effects of Foreign Trade on the Distribution of Income. *Economisk Tidskrift* **21**, 1–32.
- J.F. HELLIWELL (1996). Do National Borders Matter for Québec’s Trade? *Canadian Journal of Economics* **29**, 507–522.
- E. HELPMAN (1998). The Size of Regions. In *Topics in Public Economics*, D. PINES, E. SADKA & I. ZILCHA, editors, 33–54. Cambridge University Press, Cambridge.
- A. O. HIRSCHMAN (1958). *The Strategy of Economic Development*. Yale University Press, New Haven (CT).
- D. HUMMELS (1999). *Have International Transportation Costs Declined?* Mimeo, University of Chicago, West Lafayette (IN). Internet: <http://www.mgmt.purdue.edu/faculty/hummelsd>.
- IRPUD (2001). *European Transport Networks*. Institute of Spatial Planning, Dortmund. Internet: http://irpud.raumplanung.uni-dortmund.de/irpud/pro/ten/ten_e.htm.
- W. ISARD (1954). Location Theory and Trade Theory: Short-Run Analysis. *Quarterly Journal of Economics* **68**, 305–320.

- A.B. JAFFE, M. TRAJTENBERG & R. HENDERSON (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics* **108**, 577–598.
- M. KAPOOR, H. KELEJIAN & I.R. PRUCHA (2007). Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics* **140**, 97–130.
- H. KELEJIAN & I.R. PRUCHA (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* **17**, 99–121.
- H. KELEJIAN & I.R. PRUCHA (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* **40**, 509–533.
- W. KELLER (2002). Geographic Localization of International Technology Diffusion. *American Economic Review* **92**, 120–142.
- J.R. KENWORTHY & F.B. LAUBE (1999). An International Sourcebook of Automobile Dependence in Cities, 1960-1990. *Transportation Research A* **33**, 691–723.
- S. KIM (1995). Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987. *Quarterly Journal of Economics* **110**, 881–908.
- T. KNAAP (2004). *Models of Economic Geography: Dynamics, Estimation and Policy Evaluation*. Dissertation, Labyrinth, Ridderkerk.
- T. KNAAP (2006). Trade, Location, and Wages in the United States. *Journal of Regional Science and Urban Economics* **36**, 595–612.
- P. KRUGMAN (1979). Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics* **9**, 469–479.
- P. KRUGMAN (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review* **70**, 950–959.
- P. KRUGMAN (1991). Increasing Returns and Economic Geography. *Journal of Political Economy* **99**, 484–499.
- P. KRUGMAN & A.J. VENABLES (1995). Globalisation and the Inequality of Nations. *Quarterly Journal of Economics* **110**, 857–880.

- H. LAI & D. TREFLER (2002). *The Gains from Trade with Monopolistic Competition: Specification, Estimation, and Mis-Specification*. NBER Working Paper 9169, Washington.
- J.P. LESAGE (1999). *The Theory and Practice of Spatial Econometrics*. Department of Economics, University of Toledo. Internet: <http://www.spatial-econometrics.com/html/sbook.pdf>.
- J.P. LESAGE & R.K. PACE (2004). *Arc Mat, a Matlab Toolbox for Using ArcView Shape Files for Spatial Econometrics and Statistics*. University of Toledo, Louisiana State University. Internet: http://www.spatial-econometrics.com/html/arc_map.pdf.
- N. LIMA & A.J. VENABLES (2001). Infrastructure, Geographical Disadvantage, Transport Costs and Trade. *World Bank Economic Review* **15**, 451–479.
- P. MARTIN (1998). Can Regional Policies Affect Growth and Geography in Europe? *European Economic Review* **46**, 757–774.
- J. MCCALLUM (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *American Economic Review* **85**, 615–623.
- P. MCCANN (2005). Transport Costs and New Economic Geography. *Journal of Economic Geography* **5**, 305–318.
- P. MCCULLAGH & J.A. NELDER (1989). *Generalized Linear Models (Second Edition)*. Chapman and Hall, London.
- G. MION (2004). Spatial Externalities and Empirical Analysis: The Case of Italy. *Journal of Urban Economics* **56**, 97–118.
- P.A.P. MORAN (1950). Notes on Continuous Stochastic Phenomena. *Biometrika* **3**, 17–23.
- G. MYRDAL (1957). *Economic Theory and Under-Developed Regions*. Duckworth, London.
- J.P. NEARY (2001). Of Hype and Hyperbolas: Introducing New Economic Geography. *Journal of Economic Literature* **39**, 536–561.
- A. NIEBUHR (2004). *Market Access and Regional Disparities*. HWWA Discussion Paper 269, Hamburg.

- V. NITSCH (2000). National Borders and International Trade: Evidence from the European Union. *Canadian Journal of Economics* **22**, 1091–1105.
- B.G. OHLIN (1933). *Interregional and International Trade*. Harvard University Press, Cambridge (MA).
- H.G. OVERMAN, S. REDDING & A.J. VENABLES (2003). The Economic Geography of Trade, Production and Income: A Survey of Empirics. In *Handbook of International Trade*, E.K. CHOI & J. HARRIGAN, editors, 353–387. Blackwell, Oxford.
- D.C. PARSLEY & S.J. WEI (2001). Explaining the Border Effect: The Role of Exchange Rate Variability, Shipping Costs, and Geography. *Journal of International Economics* **55**, 87–105.
- D. PUGA (1998). The Rise and Fall of Regional Disparities. *European Economic Review* **43**, 303–343.
- J.E. RAUCH (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics* **34**, 380–400.
- INTERNATIONAL BANK FOR RECONSTRUCTION & DEVELOPMENT (Various Years). *World Development Indicators*. World Bank, Washington.
- S. REDDING & D. STURM (2005). *The Costs of Remoteness: Evidence from German Division and Reunification*. CEPR Discussion Papers 5015, London.
- S. REDDING & A.J. VENABLES (2004). Economic Geography and International Inequality. *Journal of International Economics* **62**, 53–82.
- D. RICARDO (1817). *On the Principles of Political Economy, and Taxation*. Murray, London.
- S. ROSENTHAL & W. STRANGE (2001). The Determinants of Agglomeration. *Journal of Urban Economics* **50**, 191–229.
- P.A. SAMUELSON (1952). The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent. *Economic Journal* **62**, 278–304.
- T. SCHWANEN, F.M. DIELEMAN & M. DIJST (2004). The Impact of Metropolitan Structure on Commute Behaviour in the Netherlands. *Growth and Change* **35**, 304–333.

- M.E. SPENCE (1976). Product Selection, Fixed Costs, and Monopolistic Competition. *Review of Economic Studies* **43**, 217–236.
- M. SURI, T.A. HULD & E.D. DUNLOP (2005). PVGIS: A Web-Based Solar Radiation Database for the Calculation of PV Potential in Europe. *International Journal of Sustainable Energy* **24**, 55–67.
- T. TABUCHI & J.F. THISSE (2002). Taste Heterogeneity, Labour Mobility and Economic Geography. *Journal of Development Economics* **69**, 155–177.
- J. TINBERGEN (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. Twentieth Century Fund, New York.
- F. TRIONFETTI (2001). Using Home-Biased Demand to Test Trade Theories. *Weltwirtschaftliches Archiv* **137**, 404–426.
- H.R. VARIAN (2006). *Intermediate Microeconomics: A Modern Approach (Sixth Edition)*. Norton, New York.
- A.J. VENABLES (1996). Equilibrium Location and Vertically Integrated Industries. *International Economic Review* **37**, 341–359.
- M. VERBEEK (2004). *A Guide to Modern Econometrics (Second Edition)*. John Wiley and Sons, Chichester.
- J. VINER (1950). *The Customs Union Issue*. Anderson Kramer Associates, Washington (DC).
- R.W.M. WEDDERBURN (1974). Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method. *Biometrika* **61**, 439–447.
- S.J. WEI (1996). *Intra-National Versus International Trade: How Stubborn are Nations in Global Integration?* NBER Working Paper 5531, Washington.
- H. WHITE (1980). A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* **48**, 817–838.
- J.M. WOOLDRIDGE (2001). *Econometrics of Cross Section and Panel Data*. MIT Press, Cambridge (MA).

Curriculum Vitae

Name Nils Hendrik Happich
Address Bornheimer Landstr. 37, 60316 Frankfurt am Main
EMail nils.happich@gmx.de
Date/pl. of birth 13/04/1977 in Wuppertal
Nationality German

Current Position

07/2007 - today **Commerzbank AG, Frankfurt am Main**
Specialist Public Affairs/Issue Management

Education

04/2004 - today **Christian-Albrechts-University (CAU), Kiel**
Quantitative Economics, doctoral studies
(grant by Land Schleswig-Holstein 2005, 2006)

10/1997 - 01/2004 **CAU, Kiel/Université Paris 1 (Panthéon-Sorbonne), Paris**
Economics, Diplom-Volkswirt

09/1987 - 06/1996 **Carl-Fuhlrott-Gymnasium (Secondary School), Wuppertal**
Abitur (A-levels)

Publications

N. Happich (2004). Veränderungen der europäischen Landkarte ökonomischer Aktivität in Folge der Osterweiterung: Ein Betätigungsfeld für politische Intervention? *Rissener Einblicke* 10/11, 43-59.

Additional Skills

Languages German, English, French
Software MS Office (Word, Excel, Power Point), LaTeX, MATLAB, SPSS

Kiel, July 8, 2009

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich meine Doktorarbeit "*Regional Income Differences and Market Accessibility in the Enlarged European Union: A Simultaneous Spatial CGE and Estimation Approach*" selbständig und ohne fremde Hilfe angefertigt habe und dass ich alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen meiner Arbeit, besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert habe.

Kiel, 8. Juli 2009

(Nils Happich)