

Bírálat Rappai Gábor

A MODELLEZÉS SAJÁTOSSÁGAI IDŐSORI ANOMÁLIÁK ESETÉN

MTA doktori művéről

A doktori mű az idősorok elemzésekor felmerülő statisztikai-módszertani eszközök közül tekint át négyet: a rendszertelenül megfigyelt (nemekvidisztans), az outlierekkel terhelt, a strukturális törést tartalmazó, illetve az időben aggregált idősorok modellezésének nehézségeit. A doktori értekezés fő tudományos kérdése, hogy sztenderd tesztek alkalmazva ebben a négy vizsgált kérdéskörben adódhatnak-e és milyen jellegű problémák, amelyek az ok-okozati összefüggés értelmezését megnehezítik, avagy eltorzítják. A doktori mű 6 érdemi fejezetből áll, az irodalomjegyzékkel együtt 191 oldal terjedelmű.

Az értekezés fő témaköre, azon belül a négy résztéma statisztika-módszertani szempontú vizsgálata véleményem szerint is napjainkban igen aktuálissá vált. Egyetértek a szerzővel, az adatbázisok és a matematikai-statisztikai szoftverek könnyű elérhetősége olyanok számára is lehetővé tették idősorok vizsgálatát, akik adott esetben nincsenek tisztában az egyes tesztek módszertani korlátaival. Amíg egy-két évtizede a matematikai-statisztikai szoftverek csak szűk körben voltak használatosak, és a felhasználói teamben általában volt felkészült matematikus vagy statisztikus, addig ma már pl. mesterképzésekben is oktatjuk az SPSS-t, doktori képzésekben a STATA-t stb. Ez pedig odavezetett, hogy olyanok is végeznek idősorelemzéseket, akik módszertanilag nem eléggé felkészültek és nem véletlen, hogy komolyabb folyóiratok ma már elkérlik a tanulmány adatsorait és leellenőrzik a próbákat. Tehát a nemzetközi trendek és saját tapasztalataim alapján is az értekezés fontos és aktuális kérdéseket boncolgat, egyetértek a szerző témaválasztásával és motivációjával.

A dolgozat 2. fejezete bevezeti az alapfogalmakat és modelleket, amelyekre a négy lényegi témakör épül. A fogalmi háttér bemutatása korrekt, egyedül az okság filozófiai háttérének felvázolásakor hiányolom, hogy a fontos és nagyon összetett kérdéskörből nincs eléggé kihangsúlyozva a társadalomtudományok speciális oksági megközelítése. Habár az értekezésben bemutatott példák illusztrációnak minősülnek, de többségük közgazdaságtani, ezért jó lett volna kitérni a társadalmi valóság speciális megismerési sajátosságaira. Amíg a természettudományok, kísérleti tudományok főleg determinisztikus alapon állnak, addig a társadalomtudományokban konszenzuális jellegűknél fogva többféle közmegegyezés alakult

ki, több irányzat, több paradigma próbálja leírni a társadalmi valóságot és az oksági viszonyokat. Leegyszerűsítve, a Newton által fémjelzett megközelítéstől eltér a Darwin-ra visszavezethető evolúciós megközelítés. Ezzel a kétféle alapállással is összefügg az értekezésben is jelzett közgazdaságtudományi és közgazdaságtani megközelítések eltérő alapállása és a trendek értelmezése, amíg a 'science' jellegű az operacionalizálásra helyezi a hangsúlyt és kvantitatív modellekre, addig a szoft irányzatok inkább a konceptualizálásra és kvalitatív szempontokra. Sarkítva, előbbi esetben az idősor elemzése, modellezése sokszor a vizsgálat célja, utóbbi esetben mindig csupán az egyik eszköze.

Az értekezés 3. fejezete az egyik részletesen vizsgált témakörrel, a rendszertelenül megfigyelt (nemekvidisztans) adatokból származó idősorok tesztelésekor felmerülő problémákkal, anomáliákkal foglalkozik. Áttekinti a probléma megoldására javasolt eljárás csoportokat, kiemeli előnyeiket és hátrányaikat. A dolgozatban részletesen kitér a jelölt az egyes próbák érzékenységvizsgálatára is. Szimulációval előállított idősorokat vizsgál kétféle módon, véletlenszám-generálással és szisztematikusan kihagyott értékeket tartalmazó adatsorok előállításával. Úgy vélem, hogy a jelöltnek az alapmodelleken végrehajtott szimulációi, az ezekből származó eredményei és megállapításai fontos hozzájárulások a témakör szakirodalmához, és megerősítették az ún. numerikus analízisben használatos, interpolációval történő kiegészítések gyakorlatát. Egy empirikus példán, a GDP és az export növekedése közötti kapcsolaton keresztül mutatja be az oksági viszonyok vizsgálatának problémáit. A jelölt jelzi, hogy ez csak illusztráció, de megjegyzem, hogy a termelési oldalról számolt GDP-nél a bruttó hozzáadott érték (GVA) és a tényleges export elszámolása között időbeli eltérés lehet, továbbá a GDP-t szezonálisan növelhetik a nontraded jellegű ágazatok, pl. a turizmus avagy az építőipar.

A jelölt a doktori mű 4. fejezetében az idősorokban előforduló outlierekkel foglalkozik, ami a közgazdasági idősorok elemzésénél gyakran érzékelt probléma. Az outlierok tipizálását követően az outlier-szűrésre kidolgozott statisztikai eljárások két nagy csoportját elemzi, a modell-független és a reziduális változón alapuló módszereket. Egyetérthetünk megállapításával, egy adott eloszlást feltételező modell-független technikák túl általánosak ahhoz, hogy konkrét esetekben mindig megfelelően működjenek. A speciális feltételekre épülő modellalapú eljárások már megbízhatóbbak, ezek közül a szakirodalom alapján kettőt ismertet a jelölt, a likelihood-arány tesztet és az érzékenységvizsgálatra alapozó eljárást. Fiatal kollégáival együtt kidolgozott egy új eljárást is, amely az outliernek tűnő értékeket, időpontokat egy lépésben kijelölve, dummy változók modellbe építését követően a reziduális szórás minimalizálására épít. Ezen új eljárás előnyeit a jelölt egy meggyőző példán

szemléltette. Szintén szimulációval illusztrálja az outlier-szűrés gyakorlati nehézségeit, pl. a rejtőzködő trendek kimutatását. Véleményem szerint ez a fejezet az értekezés egyik erőssége, a jelölt közreműködésével kidolgozott eljárás és a szimulációs eredmények meggyőzőek.

Az értekezés 5. fejezete a strukturális törések vizsgálatával foglalkozik. Áttekinti mind az ismert, mind az ismeretlen időpontban levő törések szakirodalmát, a leggyakrabban használt tesztek előnyeit és hátrányait. Megjegyzem, hogy a 2008-as válság felszínre hozta a strukturális törések iránti érdeklődést, pl. a gazdasági növekedés és fejlődés irodalmában a resilience modellezése, a törést követő talpra állás ok-okozati viszonyainak feltárása napjaink egyik izgalmas kutatási kérdése. Szintén szimulációval tett kísérletet a jelölt, hogy a strukturális törések egyes próbáit ellenőrizze. Egyik eredményeként kiemeli, hogy célszerű már a modellspecifikáció során tekintettel lenni a strukturális törésre, még akkor is, ha nem vagyunk biztosak abban, hogy ennek hatása szignifikáns.

A doktori mű 6. fejezete az aggregált idősorok modellezésével foglalkozik, amely szintén gyakori, pl. tőzsdei ármozgások elemzésekor. A jelölt a szakirodalom alapján áttekinti az alapprobléma két esetét, az állapot és a tartam jellegű idősorokat. Kiemelten foglalkozik a stacionárius folyamatok időbeli aggregálásával és a sztochasztikus trendeket tartalmazó, továbbá a kointegrált idősorokkal. A korábbi fejezetek logikájához hasonlóan szimulációkkal, öt különböző aggregálási rend alapján vizsgálta az aggregálás hatását az idősorokra, a trend megítélésének, az ok-okozati összefüggések elfedése, továbbá a hamis okság okozta problémákat. Az eredmények közül kiemelem, hogy az oksági viszonyok elfedésének valószínűsége az aggregálás rendjével párhuzamosan növekszik, a havi bontásról évesre, vagy napi bontásról havira történő átállás jelentősen befolyásolhatja az elemzések végkövetkeztetéseit.

Az értekezés az etikus idősor-modellezéssel kapcsolatos néhány alapvető felvetéssel zárul, ami szintén napjaink egyik fontos kérdésköre, nemcsak az egyes hivatalok kimutatásaiban, hanem a piaci szereplők elemzéseiben is lehetnek olyan lépések és megállapítások, amelyet egy korrekt elemző nem követ el. Két jelenséget emel ki, az egyik a vizsgálandó jelenségek pontos definiálása, a másik a változók pontos tartalmának megadása, főleg azonos elnevezésű változók esetében. Tapasztalataim alapján ezek valós problémák, és teljesen egyetérthetünk az értekezés 181. oldalán található 4 intellemmel.

Az értekezés nyelvezete és stílusa élvezetes, az egyes fejezetek felépítése arányos, a jelölésrendszerek konzekvensek. Formailag is igényes a doktori mű, megfelel az általános elvárásoknak. Külön kiemelem, hogy az értekezés olvasása során szembetűnő a jelölt

törekvése a világos és egyértelmű megfogalmazásra, az olvasó kiszolgálására, amiben a jelölt több évtizedes sikeres egyetemi oktatói tapasztalatai is visszaköszönnek.

A fentieket összegezve Rappai Gábor doktori munkájának tudományos eredményeit elegendőnek tartom az MTA doktora cím megszerzéséhez, a nyilvános védés kitűzését javaslom.

Szeged, 2017. május 28.

Prof. dr. Lengyel Imre
az MTA doktora

Szegedi Tudományegyetem

Bírálat Rappai Gábor

A MODELLEZÉS SAJÁTOSSÁGAI IDŐSORI ANOMÁLIÁK ESETÉN

MTA doktori művéről

A doktori mű az idősorok elemzésekor felmerülő statisztikai-módszertani eszközök közül tekint át négyet: a rendszertelenül megfigyelt (nemekvidisztans), az outlierekkel terhelt, a strukturális törést tartalmazó, illetve az időben aggregált idősorok modellezésének nehézségeit. A doktori értekezés fő tudományos kérdése, hogy sztenderd tesztek alkalmazva ebben a négy vizsgált kérdéskörben adódhatnak-e és milyen jellegű problémák, amelyek az ok-okozati összefüggés értelmezését megnehezítik, avagy eltorzítják. A doktori mű 6 érdemi fejezetből áll, az irodalomjegyzékkel együtt 191 oldal terjedelmű.

Az értekezés fő témaköre, azon belül a négy résztéma statisztika-módszertani szempontú vizsgálata véleményem szerint is napjainkban igen aktuálissá vált. Egyetértek a szerzővel, az adatbázisok és a matematikai-statisztikai szoftverek könnyű elérhetősége olyanok számára is lehetővé tették idősorok vizsgálatát, akik adott esetben nincsenek tisztában az egyes tesztek módszertani korlátaival. Amíg egy-két évtizede a matematikai-statisztikai szoftverek csak szűk körben voltak használatosak, és a felhasználói teamben általában volt felkészült matematikus vagy statisztikus, addig ma már pl. mesterképzésekben is oktatjuk az SPSS-t, doktori képzésekben a STATA-t stb. Ez pedig odavezetett, hogy olyanok is végeznek idősorelemzéseket, akik módszertanilag nem eléggé felkészültek és nem véletlen, hogy komolyabb folyóiratok ma már elkérlik a tanulmány adatsorait és leellenőrzik a próbákat. Tehát a nemzetközi trendek és saját tapasztalataim alapján is az értekezés fontos és aktuális kérdéseket boncolgat, egyetértek a szerző témaválasztásával és motivációjával.

A dolgozat 2. fejezete bevezeti az alapfogalmakat és modelleket, amelyekre a négy lényegi témakör épül. A fogalmi háttér bemutatása korrekt, egyedül az okság filozófiai háttérének felvázolásakor hiányolom, hogy a fontos és nagyon összetett kérdéskörből nincs eléggé kihangsúlyozva a társadalomtudományok speciális oksági megközelítése. Habár az értekezésben bemutatott példák illusztrációnak minősülnek, de többségük közgazdaságtani, ezért jó lett volna kitérni a társadalmi valóság speciális megismerési sajátosságaira. Amíg a természettudományok, kísérleti tudományok főleg determinisztikus alapon állnak, addig a társadalomtudományokban konszenzuális jellegűknél fogva többféle közmegegyezés alakult

ki, több irányzat, több paradigma próbálja leírni a társadalmi valóságot és az oksági viszonyokat. Leegyszerűsítve, a Newton által fémjelzett megközelítéstől eltér a Darwin-ra visszavezethető evolúciós megközelítés. Ezzel a kétféle alapállással is összefügg az értekezésben is jelzett közgazdaságtudományi és közgazdaságtani megközelítések eltérő alapállása és a trendek értelmezése, amíg a 'science' jellegű az operacionalizálásra helyezi a hangsúlyt és kvantitatív modellekre, addig a szoft irányzatok inkább a konceptualizálásra és kvalitatív szempontokra. Sarkítva, előbbi esetben az idősor elemzése, modellezése sokszor a vizsgálat célja, utóbbi esetben mindig csupán az egyik eszköze.

Az értekezés 3. fejezete az egyik részletesen vizsgált témakörrel, a rendszertelenül megfigyelt (nemekvidisztans) adatokból származó idősorok tesztelésekor felmerülő problémákkal, anomáliákkal foglalkozik. Áttekinti a probléma megoldására javasolt eljárás csoportokat, kiemeli előnyeiket és hátrányaikat. A dolgozatban részletesen kitér a jelölt az egyes próbák érzékenységvizsgálatára is. Szimulációval előállított idősorokat vizsgál kétféle módon, véletlenszám-generálással és szisztematikusan kihagyott értékeket tartalmazó adatsorok előállításával. Úgy vélem, hogy a jelöltnek az alapmodelleken végrehajtott szimulációi, az ezekből származó eredményei és megállapításai fontos hozzájárulások a témakör szakirodalmához, és megerősítették az ún. numerikus analízisben használatos, interpolációval történő kiegészítések gyakorlatát. Egy empirikus példán, a GDP és az export növekedése közötti kapcsolaton keresztül mutatja be az oksági viszonyok vizsgálatának problémáit. A jelölt jelzi, hogy ez csak illusztráció, de megjegyzem, hogy a termelési oldalról számolt GDP-nél a bruttó hozzáadott érték (GVA) és a tényleges export elszámolása között időbeli eltérés lehet, továbbá a GDP-t szezonálisan növelhetik a nontraded jellegű ágazatok, pl. a turizmus avagy az építőipar.

A jelölt a doktori mű 4. fejezetében az idősorokban előforduló outlierekkel foglalkozik, ami a közgazdasági idősorok elemzésénél gyakran érzékelt probléma. Az outlierok tipizálását követően az outlier-szűrésre kidolgozott statisztikai eljárások két nagy csoportját elemzi, a modell-független és a reziduális változón alapuló módszereket. Egyetérthetünk megállapításával, egy adott eloszlást feltételező modell-független technikák túl általánosak ahhoz, hogy konkrét esetekben mindig megfelelően működjenek. A speciális feltételekre épülő modellalapú eljárások már megbízhatóbbak, ezek közül a szakirodalom alapján kettőt ismertet a jelölt, a likelihood-arány tesztet és az érzékenységvizsgálatra alapozó eljárást. Fiatal kollégáival együtt kidolgozott egy új eljárást is, amely az outliernek tűnő értékeket, időpontokat egy lépésben kijelölve, dummy változók modellbe építését követően a reziduális szórás minimalizálására épít. Ezen új eljárás előnyeit a jelölt egy meggyőző példán

szemléltette. Szintén szimulációval illusztrálja az outlier-szűrés gyakorlati nehézségeit, pl. a rejtőzködő trendek kimutatását. Véleményem szerint ez a fejezet az értekezés egyik erőssége, a jelölt közreműködésével kidolgozott eljárás és a szimulációs eredmények meggyőzőek.

Az értekezés 5. fejezete a strukturális törések vizsgálatával foglalkozik. Áttekinti mind az ismert, mind az ismeretlen időpontban levő törések szakirodalmát, a leggyakrabban használt tesztek előnyeit és hátrányait. Megjegyzem, hogy a 2008-as válság felszínre hozta a strukturális törések iránti érdeklődést, pl. a gazdasági növekedés és fejlődés irodalmában a resilience modellezése, a törést követő talpra állás ok-okozati viszonyainak feltárása napjaink egyik izgalmas kutatási kérdése. Szintén szimulációval tett kísérletet a jelölt, hogy a strukturális törések egyes próbáit ellenőrizze. Egyik eredményeként kiemeli, hogy célszerű már a modellspecifikáció során tekintettel lenni a strukturális törésre, még akkor is, ha nem vagyunk biztosak abban, hogy ennek hatása szignifikáns.

A doktori mű 6. fejezete az aggregált idősorok modellezésével foglalkozik, amely szintén gyakori, pl. tőzsdei ármozgások elemzésekor. A jelölt a szakirodalom alapján áttekinti az alapprobléma két esetét, az állapot és a tartam jellegű idősorokat. Kiemelten foglalkozik a stacionárius folyamatok időbeli aggregálásával és a sztochasztikus trendeket tartalmazó, továbbá a kointegrált idősorokkal. A korábbi fejezetek logikájához hasonlóan szimulációkkal, öt különböző aggregálási rend alapján vizsgálta az aggregálás hatását az idősorokra, a trend megítélésének, az ok-okozati összefüggések elfedése, továbbá a hamis okság okozta problémákat. Az eredmények közül kiemelem, hogy az oksági viszonyok elfedésének valószínűsége az aggregálás rendjével párhuzamosan növekszik, a havi bontásról évesre, vagy napi bontásról havira történő átállás jelentősen befolyásolhatja az elemzések végkövetkeztetéseit.

Az értekezés az etikus idősor-modellezéssel kapcsolatos néhány alapvető felvetéssel zárul, ami szintén napjaink egyik fontos kérdésköre, nemcsak az egyes hivatalok kimutatásaiban, hanem a piaci szereplők elemzéseiben is lehetnek olyan lépések és megállapítások, amelyet egy korrekt elemző nem követ el. Két jelenséget emel ki, az egyik a vizsgálandó jelenségek pontos definiálása, a másik a változók pontos tartalmának megadása, főleg azonos elnevezésű változók esetében. Tapasztalataim alapján ezek valós problémák, és teljesen egyetérthetünk az értekezés 181. oldalán található 4 intellemmel.

Az értekezés nyelvezete és stílusa élvezetes, az egyes fejezetek felépítése arányos, a jelölésrendszerek konzekvensek. Formailag is igényes a doktori mű, megfelel az általános elvárásoknak. Külön kiemelem, hogy az értekezés olvasása során szembetűnő a jelölt

törekvése a világos és egyértelmű megfogalmazásra, az olvasó kiszolgálására, amiben a jelölt több évtizedes sikeres egyetemi oktatói tapasztalatai is visszaköszönnek.

A fentieket összegezve Rappai Gábor doktori munkájának tudományos eredményeit elegendőnek tartom az MTA doktora cím megszerzéséhez, a nyilvános védés kitűzését javaslom.

Szeged, 2017. május 28.

Prof. dr. Lengyel Imre
az MTA doktora

Szegedi Tudományegyetem