СПРОВЕДУВАЊЕ НА ЕМПИРИСКО ИСТРАЖУВАЊЕ[1]

Во ова поглавје дискутирани се составните делови на една успешна емпириска анализа со акцент на спроведување еден комплетен проект. Дополнително, се навраќаме на теми кои се неопходни во применетото истражување како потсетување на позначајните работи што се појавија низ книгата. Исто така, понудени се сугестии за некои прашања како своевиден начин за стимулирање на имагинацијата на читателот. Како додаток дадени се повеќе извори за економско истражување и референци за податоци.

1. Поставување прашање

Значењето на поставување специфично прашање никогаш не може да биде пренагласено. Без да се биде експлицитен за целта на вашата анализа, нема дури да можете да знаете од каде да почнете. Општата расположивост на богати датотеки со податоци прави да бидеме изложени на искушение за отпочнување со прибирање податоци базирано на „полупечени“ идеи, но тоа често е контрапродуктивно. Очигледно е дека без внимателно формулирање на вашите хипотези и типот на моделот што ќе треба да го оцените, вие ќе заборавите да приберете информации за важни променливи, ќе добиете примерок од погрешна популација или ќе приберете податоци за погрешен временски период.

Ова не значи дека вашето прашање треба да биде поставено во вакум. Особено за едносеместрален проект вие не можете да бидете претерано амбициозни. Според тоа, кога избирате проблем треба да бидете разбирливо сигурни дека бараните извори на податоци постојат и дека ќе ви овозможат да одговорите на вашето прашање во доделеното време.

Исто така, кога избирате еден проблем треба да одлучите кои области од економијата или другите општествени науки ве интересираат. На пример, доколку сте изучувале економија на трудот можеби сте се сретнале со теории што можат емпириски да бидат тестирани или пак, зависности што имаат некакво значење за политиките. Економистите на труд постојано изнаоѓаат нови променливи што можат да ги објаснат разликите во платите. Примерите ги вклучуваат квалитетот на високото образование, бројот на часови по математика и природни науки, физичката појава и тн. Истражувачите на државното и локалното јавно финансирање истражуваат како економската активност зависи од променливите на економските политики како што се даноците на имот, даноците на промет, нивото и квалитетот на услугите и тн.

Економистите кои ги проучуваат проблемите во образованието се интересираат како вложувањата во знаење влијаат врз успехот, дали завршувањето на одреден вид училишта го подобрува успехот или, определување на факторите што влијаат врз одлуката каде да бидат лоцирани приватни училишта.

Макроекономистите се интересираат за односите помеѓу различни агрегатни временски серии, како на пример за врските меѓу растот и бруто домашниот производ, меѓу растот и инвестициите во основни средства, или за влијанието на даноците врз каматните стапки.

[1] Овој текст е превод на 19-то поглавје од книгата на *Jeffrey M. Wooldridge, “Introductory Econometrics, A modern approach”*.

Исто така, постојат причини за оценување модели коишто се претежно дескриптивни. На пример, проценителите на имот користат модели за да ја оценат вредноста на куќите што во скоро време не биле предмет на продажба. Тоа вклучува регресионен модел што ја става во релација цената на куќата со нејзините карактеристики (големина, број на спални соби, број на капатила и тн.). Како проблем за еден емпириски труд, ова не е многу возбудливо бидејќи малку е веројатно дека ќе научиме нешто што изненадува и заради тоа што ваквата анализа нема очигледни импликации врз политиките. Додавањето на стапката на криминал во соседството како објаснувачка променлива би ни овозможило да определиме колку криминалот е значаен фактор за цената на куќите, а тоа од своја страна е корисно за оценување на трошоците од криминалот.

Повеќето зависности се веќе оценети со користење на макроекономски податоци што по природа се дескриптивни. На пример, агрегатната функција на штедење може да биде користена за да се оцени агрегатната гранична склоност кон штедење, како и реакцијата на штедењето од приносот на средствата (како што се каматните стапки). Ваквата анализа би можела да се направи многу поинтересна со користење податоци на временски серии за земја што во минатото имала политички немири, при што ќе се определи дали стапките на штедење опаѓаат во периодите на политичка несигурност.

Откако ќе се одлучите за областа на истражување, постојат повеќе начини за изнаоѓање специфични трудови за конкретниот проблем. Списанието *Journal of Economic Literature (JEL)* има детален систем на класификација, така што на секој труд му се доделени група од кодови за идентификација со чија помош се категоризира во определено подрачје од економијата. Ова списание, исто така, содржи листа на статии што се објавени во голем број разни стручни списанија организирани според подрачјето, па дури има и кратки апстракти од некои статии.

Особено погодни за пребарување публикувани трудови на различни теми се интернет услужните апликации како што е *EconLit*, каде што се претплатени голем број универзитети. *EconLit* им овозможува на корисниците да прават опсежно пребарување во скоро сите економски списанија по автор, предмет, клучни зборови во насловот и тн. Друга интернет апликација за пребарување е *Social Science Citation Index* што е корисна за пребарување трудови во поширок опсег на наслови од општествените науки, вклучувајќи ги популарните трудови што често пати биле цитирани во други публикувани дела.

При размислувањето за темата на трудот треба да се имаат предвид неколку работи. Прво, проблемот не мора да има пошироки импикации врз политиките за истиот да биде интересен односно, тој може да биде од локален интерес. На пример, може да ве интересира да дознаете дали членувањето во некоја секција во универзитетот може да предизвика повисок или понизок просечен успех на студентот. Ова не мора да биде од интерес за луѓето надвор од академските кругови, но најверојатно ќе ги засегне некои луѓе во рамките на универзитетот. Од друга страна, може да проучувате проблем што на почетокот е од локален интерес но, потоа добива пошироко значење како што, на пример, е определувањето на факторите што влијаат и поитиките што можат да ја спречат злоупотребата на алкохолот во универзитетските кампуси.

Второ, многу е тешко во текот на половина или цел семестар да се направи вистинско оригинално истражување со користење на макро-економските агрегати за економијата на САД. На пример, прашањето дали растот на паричната понуда, владината потрошувачка и сл. влијаат врз економскиот раст во континуитет се проучувани од страна на професионални макроекономисти. Прашањето за тоа дали приносот од залихите или останатите средства може систематски да биде предвиден со помош на расположивите информации, заради разни причини до сега е внимателно проучувано. Тоа не значи дека треба да избегнувате оценување на макроекономски или емпириски финансиски модели, затоа што дури и користењето на понови податоци може да придонесе за конструктивна дебата. Згора на тоа, понекогаш може да изнајдете нова променлива што има значаен ефект врз економските агрегати или финансиската заработувачка и тоа откритие може да биде возбудливо.

Поентата е во тоа што примерите како што се користење на неколку дополнителни години за да се оцени стандардната Филипсова крива или, пак, функцијата на агрегатна потрошувачка за економијата на САД или некоја друга поголема земја малку е веројатно дека ќе придонесе за добивање дополнителни сознанија без оглед на тоа што можат да бидат поучни за студентите. Наместо тоа, може да се користат податоци за помала земја за да се оцени статичката или динамичката Филипсова крива, да се тестира хипотезата за ефикасни пазари и тн.

На макроекономско ниво постојат голем број прашања коишто биле значително проучувани. На пример, во областа на економијата на трудот публикувани се многу трудови во кои е оценето влијанието на образованието врз заработувачката. Ова прашање сé уште се проучува бидејќи е многу важно и новите податоци, како и новите економски пристапи продолжуваат да бидат развивани. На пример, некои податоци имаат подобри приближни променливи за неопсервираните способности отколку други променливи. Во други случаи, можеме да добиеме панел податоци или податоци од природен експеримент што би ни овозможило да пристапиме кон старото прашање од друга перспектива.

Друг пример се однесува на криминолозите коишто се интересираат за ефектите на разни закони врз криминалот. Прашањето дали најстрогата казна има поинаков ефект долго било јавно дебатирано. На сличен начин, економистите се интересираат за тоа дали даноците на цигарите и алкохолот ја редуцираат нивната потрошувачка. Со обезбедувањето на податоци за повеќе години на државно ниво, можат да се креираат побогати панел податоци, а тоа може да помогне да бидат одговорени главните прашања коишто се однесуваат на политиките. Згора на тоа, од скоро време постојат иновации за борба против криминалот чија ефикасност може да биде предмет на емпириска евалуација.

Кога го формулирате вашиот проблем, корисно е вашите идеи да ги дискутирате со колегите, менторот и пријателите. При тоа, треба да бидете способни да ги убедите луѓето дека одговорот на поставеното прашање побудува интерес. (Секако дека убедливо да се одговори на прашањето претставува друг проблем, но треба да се започне со интересно прашање.) Ако некој ве запраша за вашиот труд, а вие одговорите со „Пишувам нешто за каматните стапки“, вие сте одлучиле само за општото подрачје без при тоа да го формулирате вистинскиот проблем. Во друг случај, би требало да бидете

способни да кажете, на пример, „Се интересирам на кој начин промените на инфлацијата влијаат врз краткорочните каматни стапки“ и сл.

2. Преглед на литература

Сите трудови, дури и оние што се релативно кратки треба да содржат преглед на релевантната литература. Ретки се случаите каде некој се обидува да изработи емпириски проект во област каде нема претходно публикувани трудови. Доколку пребарувате во стручни списанија или користите *on-line* пребарувачки услуги со цел да дојдете до вашиот наслов, тогаш сте на добар пат да извршите преглед на литературата. Доколку самостојно изберете наслов, како на пример „Изучување на влијанието од употребата на дрогите врз успехот во вашиот универзитет“, во тој случај ќе треба да одработите малку повеќе. Сепак, услугите за *on-line* пребарување ја прават оваа работа многу полесна, на тој начин што можете да пребарувате по клучни зборови, по зборови во насловот, по автори и тн. Потоа, можете да ги прочитате апстрактите од трудовите за да видите колку истите се релевантни за вашиот труд.

Кога ја пребарувате литературата, треба да водите сметка за сродни теми кои може да не се појават при пребарувањето кога користите клучни зборови. На пример, доколку го проучувате влијанието на користењето дроги врз платите на вработените или врз просечниот успех на студентите, вие можеби би требало да погледнете трудови за тоа како користењето алкохол влијае врз горенаведените фактори. Да се знае како да се направи темелно пребарување на литературата претставува вештина која се стекнува во текот на подолг временски период.

Истражувачите се разликуваат во поглед на тоа на кој начин прегледот на литературата треба да биде инкорпориран во трудот. Некои сакаат да имаат посебна секција насловена „Преглед на литература“, додека други сакаат да го вклучат прегледот на литературата како дел од воведот. Ова, во голема мера претставува работа на личен вкус, иако може да се каже дека еден поширок преглед на литературата заслужува своја посебна секција. Доколку вашиот труд се однесува на некој напреден курс по економетрија, во тој случај прегледот на литературата ќе биде подолг, додека во случај на пократок едносеместрален курс истиот ќе биде пократок.

3. Прибирање податоци

3.1 Одлучување за најсоодветни податоци

Прибирањето податоци за завршниот труд може да биде поучна, возбудлива, а понекогаш дури и фрустрирачка задача. Најнапред, за да одговорите на поставеното прашање ќе треба да одлучите за видот на податоците што ви се потребни. Како што беше дискутирано во воведот, податоците можат да се јават во најразлични форми. Најчесто застапени видови се податоците на попречен пресек, временските серии, комбинирани податоци и панел податоци.

Голем број прашања можат да бидат анализирани со некои од структурите на податоци што претходно беа објаснети. На пример, за да се проучи дали засиленото спроведување на законите го намалува криминалот,

може да се користат податоци на попречен пресек за повеќе градови, временска серија за еден град, или панел што се состои од податоци за едни исти градови во текот на две или повеќе години.

Одлуката за тоа каков тип на податоци да се прибираат често пати зависи од самата природа на анализата. За да се одговорат прашањата на индивидуално или семејно ниво најчесто имаме пристап до податоци на попречен пресек коишто типично се добиени по пат на анкета. Потоа, мора да се запрашаме дали може да добиеме доволен број податоци за да направиме доволно убедлива анализа. На пример, да претпоставиме дека сакаме да дознаеме дали семејствата кои штедат преку пензиските фондови и имаат одредени даночни предности, имаат помали заштеди. Со други зборови, сакаме да одговориме дали штедењето преку пензиските фондови едноставно ги истиснува другите форми на штедење? За таа цел, постојат извори на податоци како што е Анкетата за финансирање на потрошувачката која содржи информации за различни видови штедење по години за разни примероци семејства. Во тој контекст, се поставуваат повеќе прашања при користењето на ваквите податоци. Можеби најзначајно е прашањето за тоа дали постои доволна контрола врз приходите, демографските фактори, склоноста кон штедењето и сл., за да се направи разумна анализа врз принципот *ceteris paribus*. Ако ваквите податоци се единствени што ги имаме на располагање, во тој случај приморани сме со нив да правиме она што можеме.

Слични прашања се поставуваат кај податоците на попречен пресек за претпријатија, градови, држави и сл. Во најголем број случаи, не е очигледно дека ќе можеме да направиме анализа врз принципот *ceteris paribus* единствено со помош на податоци на попречен пресек. На пример, било која студија за ефектите од спроведувањето на законите врз криминалот мора да ја земи предвид ендогеноста на трошоците за спроведување на законите. Кога се користи стандарден метод на регресија, може да биде многу тешко да се направи анализа врз принципот *ceteris paribus* без оглед на тоа со колку контролни фактори располагаме.

Доколку сте ги прочитале напредните поглавја за методите што се однесуваат на панел податоци можеби знаете дека исти податоци на попречен пресек во два или повеќе временски периоди може да ни овозможи да ги контролираме фиксни ефекти во текот на времето што од своја страна би создало забуна за користење на регресија со податоци на попречен пресек. Така на пример, мошне тешко е да се добијат панел податоци за семејства или на индивидуално ниво, и покрај тоа што некои податоци веќе постојат, како што е „Панел студијата за динамиката на приходите“ и истите можат да се користат на мошне убедлив начин. Панел податоци, исто така, постојат и за претпријатијата. На пример, *CompuStat* и Центарот за истражување на цените поседуваат голем број панел податоци што се однесуваат на финансиските информации за претпријатијата. Во случај на поголеми единки како што се училишта, градови, општини и сл. панел податоците полесно се обезбедуваат со оглед на тоа што тие не исчезнуваат во текот на времето и владините институции се задолжени секоја година да прибираат информации за исти променливи. На пример, Федералното истражно биро (FBI) прибира и објавува детални податоци за стапките на криминал по градови.

Податоците можат да се јават во најразличен облик. Некои од податоците, особено историските, се достапни само во печатена форма. Во случај на помал број податоци, внесувањето на податоците од печатениот извор претставува релативно лесна работа. Понекогаш статиите се публикуваат врз база на мал број податоци, особено кога се применуваат временски серии. Истите можат да бидат користени во емпириското истражување преку додавање опсервации само за последните години.

Голем број податоци се достапни во електронска форма, така што со нив е многу полесно да се работи. Од неодамна голем број податоци се складираат во електронска форма, а потоа владините институции ги продаваат во комерцијални цели на начин како што тоа го прават приватните претпријатија. Од друга страна, авторите на трудови често пати, на своја волја ги доставуваат нивните податоци по електронски пат.

Денес, на интернет можат да се најдат голем број *on-line* бази на податоци и истите се достапни за крајните корисници. Од неодмана креирани се бројни интернет страници што содржат економски и други видови податоци. Други интернет страници пак, содржат линкови до податоци што често можат да бидат предмет на интерес за економистите. Во основа, пребарувањето податоци преку интернет е доста лесно, а во иднина ќе стане уште попогодно.

3.2 Внесување и складирање на податоците

Откако сте одлучиле за типот на податоците и сте го лоцирале изворот на податоци, ќе мора истите да ги претворите во облик за користење. Доколку податоците се во електронски облик, тие веќе се во форма која овозможува натамошна обработка. Најфлексибилен формат на податоци во електронски облик е т.н. стандардна текстуална датотека (ASCII). Сите статистички и економетриски софтверски пакети овозможуваат сировите податоци да бидат складирани во овој формат. Вообичаено, доколку датотеката е уредно структуирана, директно ќе биде прочитана со било кој економетриски пакет. Датотеките со податоци што се користени во оваа книга нудат повеќе примери за тоа како вообичаено се складираат податоците на попречен пресек, временските серии, комбинираните и панел податоците. Како општо правило, податоците би требало да бидат во табеларен облик, каде секоја опсервација претставува посебен ред, додека колоните претставуваат различни променливи. Понекогаш, може да се сретнете со податоци каде секоја колона претставува опсервација, а секој ред посебна променлива. Тоа не е идеално решение, но повеќето софтверски пакети дозволуваат податоците да бидат прочитани во овој облик, а потоа да бидат трансформирани. Важно е да се знае како податоците се организирани, пред истите да бидат прочитани во економетрискиот пакет.

За податоците на временски серии постои само еден разумен начин како податоците да се внесат и зачуваат. Имено, хронолошки најраниот временски период ќе претставува прва опсервација, додека најскорешниот ќе претставува последна опсервација. Често пати, корисно е да се вклучат променливи што ќе укажуваат на годините а ако е релевантно и на кварталите и месеците. На тој начин подоцна е олеснето оценувањето на различните модели, вклучувајќи ги сезонските фактори и прекини во различни временски периоди. За комбинираните податоци најдобро е податоците на попречен пресек за првата

година да го пополнат првиот блок опсервации, потоа податоците на попречен пресек за втората година и т.н. Оваа организација не е суштинска, но многу е важно да постои променлива која ќе укажува на годината која се однесува на секоја опсервација.

За панел податоците, како што претходно беше дискутирано, најдобро е сите години за секоја опсервација на попречен пресек да бидат соседни и подредени по хронолошки редослед. За панел податоците, важно е да биде вклучено единствено обележје за секоја единка на попречен пресек паралелно со променливата што се однесува на годината.

Доколку податоците ги добиете во печатена форма, постојат повеќе можности истите да бидат внесени во компјутер. Така на пример, можете да креирате текстуална датотека со користење на стандарден едитор за текст. При тоа, потребно е секоја опсервација да започнува во нов ред и секој ред да содржи ист редослед на променливите. Поконкретно, секој ред треба да има ист број на внесување и променливите да бидат раздвоени со најмалку едно празно место. Понекогаш, некој друг начин на одделување, на пример со запирка, може да биде подобар, но тоа зависи од софтверот што го користите. Доколку за некои променливи недостасуваат податоци, ќе мора да одлучите на кој начин истите ќе бидат означени. Едноставниот начин на оставање празно поле понекогаш не функционира. Некои софтверски пакети како симбол за податок што недостасува ја прифаќаат точката. Други истражувачи преферираат да користат број, како на пример некоја невозможна вредност за конкретната променлива со цел да ги означат податоците што недостасуваат. Сепак, како што ќе биде подолу дискутирано доколку не сте претпазливи, ваквиот начин на означување може да биде опасен.

Доколку располагате со ненумерички податоци, на пример доколку сакате да вклучите имиња на лица или градови, во тој случај ќе треба да проверите кој е најдобриот начин за внесување на таквите податоци во економетрискиот пакет што ќе го користите. Вообичаено, текстуалните податоци се ставаат во единечни или двојни наводници. Во спротивно, текстуалната датотека може да содржи неприлагодено форматирање, што од своја страна бара посебна програма за да биде прочитана. Во тој случај, повторно, деталите ќе треба да ги проверите за конкретниот економетриски пакет.

Друг општо прифатлив начин на внесување на податоците е во форма на табела во *Excel*. Ваквиот начин има повеќе предности во споредба со текстуална датотека. Прво, помалку е возможно некои податоци да бидат прочитани заедно ако биде испуштено празното место, бидејќи секоја опсервација за секоја променлива е сместена во посебна келија. Второ, податоците внесени во *Excel* се погодни за манипулација како што е нивното подредување, пресметување средна вредност и сл. Оваа предност е помалку важна доколку користите софтверски пакет што дозволува софистициран менаџмент со податоците како што се *Eviews* и *Stata*. Доколку податоците првобитно се внесени во *Excel*, во тој случај истите треба да бидат експортирани во облик што ќе може да биде прочитан од економетрискиот пакет. Вообичаено, тоа може директно да се направи на сличен начин како што податоците од *Excel* можат да се експортираат во текстуални податоци од различен формат.

Трета алтернатива е податоците директно да бидат внесени во економетрискиот пакет. Тоа го одбегнува користењето на едитор за текст или *Excel*, но сепак не би било згодно доколку не можете директно да се движите од една до друга опсервација со цел да правите корекции или додавања нови податоци.

Податоците што се превземени од интернет можат да бидат во најразличен облик. Често пати, податоците се јавуваат како текстуални датотеки, но можат да бидат користени и други конвенции за одделување на променливите што особено се однесува на панел податоците. Некои податоци можат да се најдат и во облик на *Excel*-ови датотеки кои потоа можат да бидат прочитани во соодветните економетриски пакети.

3.3 Проверка, прочистување и сумирање на податоците

Кај едно емпириско истражување многу е важно што е можно подобро да се запознаете со податоците што ќе ги користите. Доколку самите ги внесувате податоците, тогаш ќе бидете приморани да дознаете сé за нив. Но, доколку податоците ги добиете од некој надворешен извор, ќе треба да одвоите време за да ја разберете нивната структура и карактеристики. Дури и податоците што масовно се користат и добро се документирани можат да содржат неисправности. Доколку користите податоци што се добиени од страна на автор на труд, мора да бидете свесни дека правилата што се користени за конструирање на податоците можат да бидат заборавени.

Претходно беше направен преглед на стандардните начини на кој можат да бидат зачувани различните типови податоци. Исто така, треба да се знае на кој начин се кодирани вредностите што недостасуваат. Најдобро е вредностите што недостасуваат да бидат означени со некој ненумерички знак како што е точката. Доколку, за кодирање на вредностите што недостасуваат се користи број како што е „999“ или „-1“, во тој случај треба да се биде многу претпазлив при пресметувањето на било какви статистики. Економетрискиот пакет што го користите сигурно нема да знае дека некој број навистина претставува вредност што недостасува. Многу е поверојатно дека таквите опсервации ќе бидат употребени како да се валидни, а тоа може да доведе до погрешни резултати. Најдобар начин е за секој нумерички код што се користи за вредностите што недостасуваат да се избере некој друг симбол (како на пример точка) којшто не може да биде погрешно заменет со вистинските податоци.

Исто така, мора да се запознаете со природата на променливите во датотеката и да дадете одговор на повеќе прашања. Кои од променливите се во бинарна форма а кои, пак, се споредбени (како на пример кредитниот рејтинг)? Во кои единици мерки се изразени променливите? На пример, дали монетарните податоци се изразени во долари, милиони долари и сл.? Дали променливите претставуваат стапки, како што се стапката на проодност во училиштата, стапката на инфлација, каматната стапка и сл. мерени како процент или соднос? Особено за податоците на временските серии, важно е да се знае дали монетарните вредности се во номинални или реални парични единици. Доколку податоците се изразени во реална смисла, тогаш треба да се знае која година е земена како базична.

Доколку податоците ги добиете од некој автор, тогаш некои променливи можеби веќе се трансформирани. На пример, понекогаш во самите датотеки се приложени само логаритамските вредности на променливите (како што се платите).

Откривањето грешки во податоците е неопходно со цел да се зачува интегритетот на анализата. Секогаш е корисно да се најдат минималните, максималните, средните вредности и стандардните девијации на сите или пак на најзначајните променливи во анализата. На пример, доколку откриете дека минималната вредност на образованието во вашиот примерок изнесува -99 како ниво на образование, тогаш знаете дека недостасува најмалку една опсервација за нивото на образование. Доколку по натамошната проверка пронајдите дека повеќе опсервации имаат ниво на образование -99, тогаш ќе бидете сигурни дека станува збор за код на вредност што недостасува за образованието. Или на пример, доколку за некоја стапка добиете просечна вредност 0,632, тогаш ќе знаете дека таа е мерена како пропорција а не како процент. Во тој случај, доколку максималната вредност изнесува над единица, очигледно е дека станува збор за печатна грешка. Не е ретко да се најдат датотеки со податоци каде повеќето од податоците за стапки се внесени како проценти, но некои се внесени како сооднос. Ваквите грешки на кодирањето на податоците тешко се откриваат, но важно е да се направи обид.

Исто така, мора да бидеме претпазливи кога се користат податоци на временски серии. Доколку податоците се на месечно или квартално ниво, мора да знаеме дали тие се сезонски прилагодени. Трансформирањето на податоците, исто така, бара особена внимателност. Да претпоставиме дека располагаме со месечни податоци и сакаме да креираме променлива за промената од еден месец во друг. За да го сториме тоа, неопходно е да бидеме сигурни дека податоците се подредени во хронолошки редослед од најраниот до најдоцниот период. Доколку, заради некои причини тоа не е случај, тогаш одземањето ќе резултира со неупотребливост на податоците. За да бидеме сигурни дека податоците се соодветно подредени, корисно е да имаме индикатор на временскиот период. За годишните податоци доволно е да се знае годината, но тогаш неопходно е да се знае дали годината е внесена со четири или две цифри (на пример 1998 наспроти 98). За месечните и кварталните податоци корисно е да имаме променливи што ќе укажуваат на месецот или кварталот. За месечните податоци може да имаме помошни променливи (11 или 12) или една променлива што укажува на месецот (од 1 до 12 или текстуална променлива што прима вредности: *jan*, *feb* и тн.)

Со помош на годишни, квартални или месечни индикатори лесно ќе може да се конструира временскиот тренд во било кој софтверски пакет. Креирањето сезонски помошни променливи е лесно доколку кварталот или месецот се јасно посочени или се знае барем месецот или кварталот на почетната опсервација.

Работата со панел податоците може да биде уште попредизвикувачка. Претходно веќе беше дискутиран методот на обични најмали квадрати за податоци на разлики, како еден општ пристап за контролирање на неопсервираните ефекти. При конструирањето на податоци на разлики треба да се биде претпазлив да не се создадат фантомски опсервации. Да претпоставиме дека имаме панел податоци за градови од 1992 до 1997 година. Дури и во случај

кога податоците се претходно хронолошки подредени за секоја единка на попречен пресек, непромисленото диференцирање ќе креира опсервација за 1992 година за сите градови со исклучок на првиот во примерокот. Таквите опсервации за секој град ќе се добијат со одземање на вредноста за 1997 од вредноста за 1992 што очигледно нема смисла. Според тоа, мора да бидеме сигурни дека годината 1992 недостасува за сите диференцирани променливи.

4. Економетриска анализа

Со оглед на обемноста на методологијата што се користи во економетријата, во оваа секција нема да биде направен преглед на економетриските методи. Сепак, можеме да дадеме некои општи упатства за видовите на проблеми што треба да бидат рагледани во една емпириска анализа. Како што беше претходно дискутирано, по изборот на проблемот што ќе биде предмет на истражување, потребно е да се обезбедат соодветни податоци. По претпоставка дека тоа е веќе направено, натаму потребно е да одлучиме за најсоодветните економетриски методи.

Доколку вашиот курс бил фокусиран на примена на методот на оценување со помош на обични најмали квадрати или повеќекратен линеарен регресионен модел, со користење било на податоци на попречен пресек или временски серии, економетрискиот пристап во голема мера е веќе определен. Тоа не претставува методолошка слабост, со оглед на тоа што методот на обични најмали квадрати сé уште е најчесто применуван економетриски метод. Сепак, неопходно е натаму да одлучите дали ќе биде потребно користење на некоја од варијантите на методот на обични најмали квадрати, како што се пондерирани најмали квадрати или најмали квадрати што се коригирани за сериската корелација кај регресиите на временски серии.

Со цел да се оправда користењето на методот на најмали квадрати ќе мора да бидете убедливи дека кај вашиот модел се задоволени клучните претпоставки на овој метод. Како што беше претходно дискутирано, првостепено значење има претпоставката дека случајната грешка не е во корелација со независните променливи. Во идеален случај, можно е да се провери за доволен број фактори дека оние што се вклучени во случајната грешка не се во некаква релација со регресорите. Особено кога на располагање имаме податоци на попречен пресек каде испитаници се лица, домаќинства или претпријатија како релевантен се поставува проблемот на соодветен избор. Во овој контекст, би можеле да бидете во состојба да тврдите дека останатите извори на ендогеност како што се грешката во мерење или симултаност не претставуваат сериозни проблеми.

При специфицирањерто на вашиот модел, меѓу другото, ќе мора да донесете одлука во поглед на функционалниот облик, што наметнува одговор на повеќе прашања. Дали некои од променливите треба да бидат логаритмирани? (Во применетата анализа, често пати одговорот е „да“) Дали некои од променливите треба да бидат изразени во нивоа или квадрирани со цел опфаќање на опаѓачките ефект? На кој начин би требало да бидат вклучени квалитативните фактори? Дали е доволно само да бидат вклучени бинарни променливи за различни атрибути или групи од испитаници или истите би требало да бидат во интеракција со квантитативните променливи?

Честа грешка, особено кај почетниците е неправилно вклучување во моделот на независни променливи кои по својата природа примаат нумерички вредности, но немаат квантитативно значење. На пример, кај податоците на индивидуално ниво, покрај платите, образованието и искуството, може да биде вклучена и променлива за професијата. Вообичаено, вредностите што ги прима оваа променлива се само произволни кодови што се доделени на различни професии. Фактот што на професијата наставник во основно образование, на пример е доделена вредноста 453, а на професијата информатичар вредност 751 е релевантна само во поглед на тоа да овозможи разликување помеѓу двете професии. Заради тоа, вклучувањето на променлива за професијата која ги прима ваквите вредности нема никаква смисла. Наместо тоа, подобро би било да се дефинираат различни помошни променливи за секоја професија или за групи од професии ако нивниот број е многу голем, а потоа помошните променливи би можеле лесно да бидат вклучени во моделот. Помалку банален проблем настанува доколку редоследна квалитативна променлива е вклучена како независна променлива. Да претпоставиме дека кај податоците за платите е вклучена променлива што ја мери сатисфакцијата од работата во граници од 1 до 7, каде 7 означува најголема сатисфакција. Доколку располагаме со доволно податоци, би дефинирале група од шест помошни променливи за сатисфакција од работата за нивоата од 2 до 7, земајќи го нивото 1 за референтна категорија. Со вклучување на шест помошни променливи за сатисфакција од работата во регресиониот модел овозможуваме сосема флексибилен однос меѓу зависната променлива и сатисфакцијата од работата. Доколку промеливата за сатисфакцијата ја вклучиме директно во моделот тоа имплицитно претпоставува дека зголемување за една единица има квантитативно значење. Без оглед на тоа што насоката на ефектот во најголем број случаи би била правилно оценета, сепак интерпретацијата на коефициентот пред редоследна променлива е потешка. Доколку една редоследна променлива прима голем број вредности тогаш можеме да дефинираме помошни променливи за рангови од вредности што ги прима таа променлива.

Понекогаш сакаме да објасниме променлива која прима вредности на редоследна скала. На пример, променливата за сатисфакција од работата може да се употреби како зависна променлива во регресиониот модел каде независни променливи се карактеристиките на работникот и работодавецот. За жал, доколку променливата за сатисфакција од работата е во оригинален облик, во тој случај коефициентите во моделот тешко ќе можат да се интерпретираат. Секој коефициент во овој случај би ја мерел промената во сатисфакцијата од работата при единечно зголемување на независните променливи. Од друга страна, некои модели како што се подреден *probit* или подреден *logit* добро се прилагодени за редоследна зависна променлива. Овие модели претставуваат значително надградување на бинарните *probit* и *logit* модели. Едноставно решение би било со претворање на редоследна зависна променлива во бинарна. На пример, можеме да дефинираме бинарна променлива што прима вредност 1 доколку сатисфакцијата од работата е најмалку 4, а 0 во останатите случаи. Сепак, креирањето бинарна променлива доведува до губење на одредена информација и наметнува произволно определување на границата.

Кај анализата на податоците на попречен пресек, второстепена но не помалку значајна работа е дали во моделот постои хетероскедастичност. Претходно беше напоменато на кој начин ќе биде третиран овој проблем.

Наједноставен начин е доколку се пресметаат статистики коишто не се осетливи на хетероскедастичност.

Како што беше потенцирано во претходните поглавја, примената на временските серии бара дополнителна грижа во однос на повеќе прашања. Дали равенката би била оценета на нивоа? Дали во тој случај е неопходен временски тренд? Дали определувањето на разлики е посоодветно? Дали е потребно да се земат во предвид сезонските варијации доколку податоците се месечни или квартални? Доколку на пример се воведува динамика преку дистрибуирани временски задоцнувања, колку периоди би требало да бидат вклучени? Во овој случај би требало да тргнете од одреден број периоди врз основа на интуиција или општа смисла, но на крај тоа би зависело од конкретните емпириски резултати.

Доколку вашиот модел потенцијално има погрешна спецификација и доколку користите метод на најмали квадрати, во тој случај ќе треба да направите анализа на функционалната форма. Дали би можеле да определите врз основа на разумни претпоставки која би била насоката на отстапувањата кај оценките?

Доколку сте го изучувале методот на инструментални променливи, ќе знаете дека истиот може да биде употребен за решавање на различни форми на ендогеност, вклучувајќи изоставување променливи, грешки во променливите и симултаност. При тоа, природна е потребата интензивно да размислувате за валидноста на инструменталните променливи што ги користите.

Добрите трудови во емпириските општествени науки содржат анализа на осетливост. Воопштено, тоа значи дека прво се оценува оригиналниот модел, а потоа истиот се менува на начини што изгледаат логично, а при тоа, очекуваме дека важните заклучоци нема да се променат. На пример, доколку за независна променлива користите мерка на консумирање алкохол, дали добивате квалитативно слични резултати доколку ја замените квантитативната мерка со помошна променлива за употребата на алкохол? Доколку бинарната помошна променлива е значајна, но квантитативната променлива не е значајна, тоа би можело да рефлектира некои неопсервирани фактори коишто се во корелација со употребата на алкохол. Сепак, тоа би требало да биде разгледувано за секој случај посебно.

Доколку некои опсервации значително се разликуваат од мнозинството во примерокот, на пример има неколку претпријатија коишто се многу поголеми од останатите во примерокот, дали резултатите драстично се менуваат доколку овие опсервации се исклучат од оценувањето? Доколку тоа е случај, ќе мора да ја промените функционалната форма и да образожите дека тие опсервации следат сосема различен модел.

Користењето на панел податоци покренува дополнителни економетриски прашања. Да претпоставиме дека сте прибрале податоци за два различни периоди. Постојат најмалку четири начини за користење на панел податоци за два периоди без употреба на инструментални променливи. Можете да ги споите опсервациите од двата периоди и да примените стандарден метод на најмали квадрати. И покрај тоа што на овој начин се зголемува примерокот во споредба со употреба на податоци на попречен пресек, сепак не е можно да се опфатат варијациите во текот на времето. Згора на тоа, случајните грешки во ваквата

равенка скоро секогаш се во сериска корелација заради ефектите што не се опсервирани. Оценката на случајниот ефект го коригира проблемот на сериска корелација и доведува до асимптотски ефикасни оценки, доколку неопсервираниот ефект има средна вредност 0 за сите вредности на независните променливи во текот на двата временски периоди.

Друг можен начин е во равенката за вториот период да се вклучи зависна променлива со временско задоцнување. Тоа претставува начин за ублажување на проблемот на изоставена променлива, со оглед на тоа дека почетниот исход од зависната променлива е фиксен. Тоа, често пати води кон слични резултати како во случајот кога се прави разлика меѓу податоците за двата периоди.

Кај панел податоците со повеќе години имаме уште една дополнителна опција покрај веќе споменатите опции. При тоа, можеме да користиме трансформација на фиксен ефект со цел да се елиминира неопсервираниот ефект. (Кај податоците со две години тоа е исто како и определување на разлики). Во претходните поглавја покажавме на кој начин техниките на инструментални променливи можат да бидат комбинирани со трансформација на панел податоци со цел претпоставките за егзогеност да бидат релаксирани. Како општо правило, добра е идејата да се применат повеќе разумни економетриски методи а потоа да се споредат резултатите од оценките. Тоа често пати овозможува да се определи која од претпоставките има веројатност да биде погрешна.

Дури и во случај кога сте многу внимателни во изборот на проблемот на истражување, градењето на моделот, прибирањето податоци и спроведувањето на економетриска анализа, сосема е можно да се добијат целосно или делимично збунувачки резултати. Кога тоа се случува, природна постапка е да се испробаат различни модели, различни техники на оценување или можеби различни подмножества податоци, сé додека резултатите не кореспондираат со она што било очекувано. Во пракса сите истражувачи испробуваат повеќе модели пред да го пронајдат „најдобриот“ модел. За жал, ваквата постапка позната како *data mining* ги нарушува претпоставките што ги направивме за економетриската анализа. Резултатите за непристрасност на оценките добиени со методот на најмали квадрати, како и t и F распределбите што ги изведовме за тестирање на хипотези, претпоставуваат дека извлечен е примерок што го следи моделот на популацијата и извршено е оценување на моделот. Оценувањето на различни варијантни модели за нашиот оригинален модел ја нарушува таа претпоставка бидејќи користиме исти податоци во потрагата по спецификација. Всушност, ги користиме резултатите од тестовите со цел да извршиме респецификација на нашиот модел. При тоа, оценките и тестовите од различните спецификации на моделот не се меѓусебно независни.

Некои начини за пребарување спецификации се веќе програмирани во стандардните софтверски пакети. Еден од најпопуларните е познат како „регресија во чекори“ (*stepwise regression*) каде различни комбинации од независни променливи се користат во повеќекратна регресиона анализа со цел да се дојде до најдбриот модел. Постојат повеќе начини на кои регресијата во чекори може да биде употребена, но истите овде нема да бидат детално прикажани. Општата идеја е било да се тргне од голем модел и да се зачуваат променливите чии p-вредности се наоѓаат под некое ниво на значајност или да се тргне од едноставен модел и да се додаваат променливи што имаат значајни

p-вредности. Понекогаш, групи од променливи се тестираат со помош на *F*-тестот. За жал, конечниот модел често пати зависи од редоследот според кој променливите биле изоставани или додавани. Згора на тоа, овој метод претставува строга форма на *data mining* и заради тоа во конечниот модел *t* и *F* статистиките тешко можат да бидат интерпретирани. Во тој контекст, може да се аргументира дека регресијата во чекори едноставно го автоматизира она што, во секој случај, истражувачите го прават кога испробуваат различни модели. Сепак, во најголем дел од апликациите од најголем интерес се една или две независни променливи, а потоа целта е да се испита кај овие променливи колку се стабилни коефициентите кога се додаваат или изоставуваат други променливи или, пак доколку се менува функционалната форма на моделот.

Во основа, резултатите од *data mining*-от можат да се инкорпорираат во нашето статистичко заклучување. Во пракса тоа е мошне тешко остварливо и ретко се прави особено во еден софистициран емпириски труд. Сепак, можеме да се обидеме да го минимизираме *data mining*-от преку избегнување да се испробуваат бројни модели или методи за оценување сè додека не се добие значаен резултат, а потоа да се прикаже само тој резултат. Доколку некоја променлива е статистички значајна само кај мал дел од оценетите модели, сосема е веројатно дека таа променлива нема ефект врз популацијата.

5. Пишување емпириски труд

Пишување труд во којшто е употребена економетриска анализа е многу предизвикувачка но, исто така може да биде и корисна работа. Еден успешен труд комбинира внимателна и убедлива анализа на податоците со добри објаснувања и излагања. Според тоа, мора да постои добро владеење со проблемот, добро разбирање на економетриските методи и солидно познавање на вештините на пишување. Затоа, не треба да бидете обесхрабрени доколку согледате дека пишувањето емпириски труд е тешко, бидејќи голем број професионални истражувачи поминале многу години учејќи како да создадат една емпириска анализа и да ги напишат резултатите во убедлива форма.

Стиловите на пишување во голема мера варираат, но многу трудови следат една иста општа шема. Следните параграфи вклучуваат идеи за наслови на заглавијата и објаснувања за тоа што би требало да содржи секоја секција. Тоа се само сугестии коишто не мора да бидат следени стриктно. Во крајната верзија од трудот на секоја секција треба да биде доделен број, вообичаено почнувајќи со број еден за воведот.

5.1 Вовед

Во воведот се формулирани основните цели на студијата и објаснета е нивната важност. Тоа, во основа бара преглед на литературата со наведување на тоа што до сега било направено и како истото може понатаму да биде подобрено (како што беше претходно дискутирано, поопширен преглед на литературата може да биде поместен во посебна секција). Презентирање на едноставни статистички податоци или графикони што откриваат наизглед контрадикторни зависности претставува корисен начин за воведување на проблемот во трудот. На пример, да претпоставиме дека пишувате труд за факторите што влијаат врз фертилитетот во една земја во развој со фокус на

нивото на образование кај жените. Еден згоден начин да се воведи проблемот би бил преку изработување табела или графикон што покажува дека фертилитетот во текот на времето опаѓа и давање кратко објаснување на кој начин би биле испитани факторите што допринеле за тоа опаѓање. Во овој стадиум, доволно е да се претпостави дека жените со повисоко образование имаат помалку деца и дека просечното ниво на образование пораснало во текот на времето.

Повеќето истражувачи во воведот сакаат да ги сумираат заклучоците од нивните трудови. Тоа може да биде корисно средство за да се задржи вниманието на читателот. На пример, можете да истакнете дека отсуството од десет часови на предавања во текот на семестарот доведува до половина помала оцена. Сепак, во ваквото сумирање не смее премногу да се навлегува бидејќи методите и податоците кои се користат за добивање на резултатите сѐ уште не се презентирани.

5.2 Концептуална (или теоретска) рамка

Оваа е секција каде што го објаснувате генералниот пристап во разработување на проблемот што сте го поставиле. Тоа може да биде формална економска теорија, но во многу случаи тоа претставува интуитивна дискусија за тоа кои концептуални проблеми се јавуваат во одговарањето на вашите поставени прашања.

На пример, да претпоставиме дека го проучувате влијанието на економските можности и строгоста на казните врз криминалното однесување. Еден можен пристап за да се објасни криминалното учество е преку специфицирање на проблемот на максимизирање на korиста, каде поединецот избира колкав дел од времето ќе потроши на легални односно илегални активности, имајќи ги предвид платите во двата типа активности, како и променливата што ја мери веројатноста и строгоста на казната за криминална активност. Користа од ваквата вежба е во тоа што сугерира кои променливи би требало да бидат вклучени во емпириската анализа и претставува водич за тоа на кој начин променливите би требало да фигурираат во економетрискиот модел.

Често пати, не постои потреба да се опишува економската теорија. Кај анализата на економските политики, вообичаено, доволно е здраворазумско расудување за специфицирање на моделот. На пример, да претпоставиме дека ве интересира да го оцените влијанието на учеството во програмата „Помош за фамилии со издржувани деца“ врз училишниот успех на детето. Оваа програма обезбедува дополнителен приход, но учеството исто така го олеснува добивањето на медицинска помош и остнати бенефиции. Потешкиот дел кај ваквата анализа се однесува на одлуката за тоа кои променливи треба да се контролираат. Во овој пример, можеме да ги контролираме семејните приходи, образованието на мајката, дали семејството живее во урбано место и други променливи. Натаму, воведувањето на променливата-индикатор што означува учество во горенаведената програма би требало да ја мери неприходната корист од учеството во прогрмата. Во овој случај, дискусијата за тоа кои фактори треба да бидат контролирани и механизмот преку којшто учеството во програмата би можело да го подобри училишниот успех претставува супституција за формалната економска теорија.

5.3 Економетриски модели и методи за оценување

Мошне корисно е трудот да содржи секција со равенки на моделите што ги оценувате а потоа ги презентирате во резултатите од истражувањето. Тоа овозможува да се фокусирате на клучните променливи и фактори што ќе ги испитувате. Претставување равенки коишто содржат случајна грешка овозможува да се дискутира дали е соодветен методот за оценување како што е на пример, методот на обични најмали квадрати.

Во оваа секција, исто така, треба да се направи дистинкција меѓу модел и метод за оценување. Моделот претставува зависности во популацијата. На пример, еден модел може да изгледа на следниов начин:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \qquad \dots (1)$$

каде што Y е зависна променлива, а X_i се независни променливи. Треба да се потенцира дека во оваа равенка над коефициентите β_i не постојат „капчиња“ бидејќи тоа е модел а не оценета равенка. Исто така, не пишуваме конкретни вредности што ги примаат коефициентите, бидејќи истите не се познати (и никогаш нема да бидат познати затоа што се однесуваат на целата популација). Коефициентите ќе бидат оценети подоцна во истражувањето. Во оваа секција не е потребно презентирање на очекувањата за вашите емпириски резултати. Со други зборови, не е препорачливо да започнете со еден генерален модел, а потоа да наведете дека сте изоставиле некои променливи бидејќи не биле значајни. Ваквата дискусија треба да биде оставена за секцијата со резултатите.

Еден модел на временска серија би можел да изгледа на следниов начин:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1t-1} + \beta_3 X_{2t} + \beta_4 X_{2t-1} + \beta_5 X_{3t} + u_t \qquad \dots (2)$$

каде што индексот t се користи со цел да се потенцира динамиката во равенката (во овој случај за променливите X_1 и X_2 земен е предвид и ефектот на временско задоцнување).

По специфицирањето на моделот, згодно е да се дискутира методот за оценување. Во најголем број случаи тоа би бил методот на обични најмали квадрати (OLS) но, на пример во равенките на временски серии може да се користи и методот на генерализирани најмали квадрати со цел да се изврши коригирање на сериската корелација. Сепак, методот за оценување на моделот сосема се разликува од самиот модел. Во тој контекст, не би имало значење да се зборува за „OLS модел“, бидејќи обичните најмали квадрати се само метод за оценување. Постојат голем број начини за оценување на моделот, а вие ќе треба да објасните зошто е оправдан избраниот метод во конкретниот случај.

Секоја претпоставка што се применува за конкретниот економетриски модел којшто е предмет на оценка треба да биде разбирливо дискутирана. На пример, доколку предмет на истражување е квалитетот на училиштата, во анализата централен ќе биде проблемот на мерење на квалитетот на училиштата. Се поставува прашање дали тоа би бил процентот на проодност, соодносот меѓу бројот на ученици и наставници, просечното ниво на образование на наставниците или, пак, некоја комбинација од сите овие или некои други показатели?

Во поглед на функционалната форма секогаш ќе треба да се направат претпоставки дали е веќе понуден некаков теоретски модел. Како што е познато, моделите со константна еластичност и полуеластичност доста се експлоатирани бидејќи коефициентите можат лесно да бидат интерпретирани (како процентуални ефекти). Сепак, не постојат цврсти правила за тоа како функционалната форма треба да биде избрана. Во тој контекст, не постои потреба за опширна дискусија за функционалната форма но, корисно е да се напомене дали ќе оценувате еластичност или полуеластичност. На пример, доколку го оценувате ефектот на некоја променлива врз платите, зависната променлива скоро секогаш ќе биде во логаритамски облик, и тоа ќе треба да го вклучите во равенката од самиот почеток. При тоа, не треба да се презентираат сите можни варијации кај функционалната форма што подоцна ќе бидат прикажани во секцијата на резултатите.

Често пати, податоците во економетриските анализи се на ниво на градови или региони. Во тој случај резултатите можат да се претставуваат на агрегатно ниво или по глава на жител. Доколку моделот вклучува податоци на агрегатно ниво, бројот на жители може да биде еден од факторите (независни променливи) што влијае врз зависната променлива. Во случај кога сакаме резултатите да бидат прикажани по глава на жител, сите променливи ќе треба да бидат поделени со бројот на жители во соодветниот град, односно регион.

Доколку користите понапредна техника за оценување, како на пример најмали квадрати во две фази, ќе треба да објасните зошто го правите тоа. Доколку користите метод на инструментални променливи, ќе треба да понудите внимателна дискусија за валидноста во изборот на инструментална променлива како ендогена независна променлива. Како што претходно беше напоменато, за да една променлива биде третирана како добра инструментална променлива треба да бидат исполнети две барања. Прво, треба да биде егзогена и како таква да биде отстранета равенката која е предмет на интересирање (структурна равенка). Тоа е нешто што мора да биде претпоставено. Второ, потребно е да постои парцијална корелација со ендогените независни променливи. Тоа е нешто што може да биде тестирано.

Исто така, кога се користат панел податоци во трудот би требало да водите сметка за проблемот на изоставени променливи (или изоставена хетерогеност). Повторно, тоа лесно се опишува со пишување на една или две равенки. При тоа, корисно е да се покаже како се разликуваат равенките во текот на времето со цел да се отстранат неопсервираните константни фактори. Тоа доведува до равенка што може да биде оценета со помош на методот на обични најмали квадрати. Доколку наместо тоа се користат оценки со фиксни ефекти, тоа ќе треба да биде експлицитно назначено. Како едноставен пример, да претпоставиме дека сакате да тестирате дали повисоките даноци во земјата ја редуцираат економската активност мерена преку обемот на производство по глава на жител. Во моделот ќе треба да бидат вклучени помошни променливи за поодделни години, променлива за висината на даноците за секоја земја по години, како и други променливи коишто се однесуваат на факторите што се менуваат во текот на времето вклучувајќи ги трошоците за занимавање со бизнис, продуктивноста на работниците и тн.

5.4 Податоци

Неопходно е секогаш да имате секција во која внимателно се опишани податоците што се користат во емпириското оценување. Ова особено е значајно доколку вашите податоци се нестандардни или пак не биле пошироко користени од страна на други истражувачи. Во тој контекст, потребно е да се понудат доволно информации за да читателот во основа може да ги добие податоците и да ја повтори вашата анализа. Особено, сите применливи јавни извори на податоци треба да бидат вклучени во референците, а кратките серии податоци можат да бидат излистани во прилог. Доколку сте направиле ваше сопствено анкетно истражување со цел да приберете податоци, тогаш во прилог треба да се презентира копија од прашалникот.

Паралелно со дискусијата за изворите на податоци, осигурете се во поглед на единиците мерки за секоја од променливите (на пример, дали приходите се мерени во стотици или илјада долари?). Вклучување на табела со дефиниции за променливите е мошне корисно за читателот. Имињата во табелата би требало да кореспондираат со имињата користени за опишување на економетриските резултати во наредната секција.

Исто така, мошне информативно е за секоја од променливите да се прикаже табела со сумарни статистики, како што се минималните и максималните вредности, аритметичките средини и стандардните девијации. Ваквата табела го олеснува интерпретирањето на оценетите коефициенти во следната секција и ги потенцира единиците мерки на променливите. За бинарните променливи единствена потербна сумарна статистика е соодносот на единици во примерокот (што е исто со аритметичката средина на примерокот). За променливите што пројавуваат тренд, помалку интересни се сумарните статистики како што е аритметичката средина. Во тој случај, често пати корисно е да се пресмета просечната стапка на раст на променливата за целокупниот период на примерокот.

Секогаш ќе треба јасно да биде назначен бројот на опсервации со кој располагате. За податоците на временски серии треба да ги идентификувате годините што ги користите во анализата, вклучувајќи и опис на некој специфичен период во историјата (како на пример, Втората светска војна). Доколку користите панел податоци осигурете се за секоја година да го прикажете бројот на единици на попречен пресек (лица, населени места и сл.).

5.5 Резултати

Секцијата на резултатите треба да ги вклучи оценките на моделите што се формулирани во секцијата на моделите. При тоа, можете да се започне со мошне едноставна анализа. На пример, да претпоставиме дека процентот на ученици од завршна година што се запишуваат на факултет се користи како мерка за квалитет на училиштето што го посетува едно лице. Во тој случај равенката која би се оценувала е следна:

$$\log(Y) = \beta_0 + \beta_1 X + u$$

каде што Y е висина на платата, а X е процент на ученици од завршна година што се запишуваат на факултет. Секако дека ваквата формулација не овозможува проверка на повеќе други фактори коишто би можеле да ја

детерминираат висината на платата и би биле во корелација со независната променлива. Сепак, едноставната анализа може да го воведе читателот во натамошна пософистицирана анализа и да ја открие важноста за проверка на другите фактори.

Доколку се оценети само неколку равенки, резултатите можат да се претставaт во форма на равенка со прикажување во загради на стандардните грешки под оценетите коефициенти. Доколку вашиот модел има повеќе независни променливи и доколку презентирате неколку варијации на основниот модел, подобро е резултатите да се претставaт во табеларен облик, отколку во облик на равенка. Во најголем број случаи би имале барем една табела којашто секогаш треба да ги содржи коефициентите на детерминација и бројот на опсервации за секоја равенка. Исто така, може да бидат прикажани и други статистички показатели како што е прилагодениот коефициент на детерминација.

Најважно е да се дискутираат силните страни на вашите емпириски резултати. При тоа, треба да се даде одговор на повеќе прашања. Дали коефициентите ги имаат очекуваните знаци? Дали тие се статистички значајни? Доколку некој коефициент е статистички значаен но, има знак што е спротивен на претходните очекувања зошто тоа би можело да е исправно? Тоа би можело да открива проблем со податоците или методот на оценување (На пример, методот на обични најмали квадрати може да биде несоодветен како резултат на проблемот на изоставена променлива).

Исто така, треба да бидете сигурни дека сте ги објасниле вредностите на коефициентите пред значајните независни променливи. Често пати, централно место во истражувањето имаат една или две променливи на политиките. Нивните знаци, вредности и статистичката значајност би требало да бидат детално третирани. При тоа, запомнете да направите дистинкција меѓу економска и статистичка значајност. Доколку *t* статистиката е мала, треба да се наведе дали тоа се должи на фактот дека коефициентот практично е многу мал или дека неговата стандардна грешка е голема?

Во продожение на дискусијата на оценките од општиот модел, можете да понудите интересни посебни испитувања, особено оние што се потребни за да се тестираат некои повеќекратни хипотези. На пример, во истражувањето за разликите во платите меѓу различните индустриски гранки можете да прикажете равенка без помошни променливи за индустриските гранки. Тоа му овозможува на читателот лесно да тестира дали разликите се статистички значајни (преку користење на коефициентот на детерминација или *F*-тестот). При тоа, не грижете се многу за испуштање на различни променливи со цел да се изнајде „најдобрата“ комбинација од независни променливи. Како што претходно спомнавме, тоа е тешка задача која не е сосема дефинирана. Тоа е значајно само во случај кога елиминирање на група променливи значително ги менува вредностите и/или значајноста на коефициентите што се предмет на интерес. Испуштање на група променливи како што е квадратен член или член на интеракција со цел поедноставување на моделот, може да биде оправдано со помош на *F*-тестот.

Доколку сте користеле најмалку два различни методи како на пример методот на обични најмали квадрати и најмали квадрати во две фази, или нивоа и разлики кај временските серии во тој случај би требало да ги коментирате сите

критични разлики. Особено, доколку методот на обични најмали квадрати дава резултати што се спротивни на очекувањата, дали користењето на методот на најмали квадрати во две фази или панел податоци ги подобруваат оценките?

5.6 Заклучоци

Ова би било кратка секција во која се сумира она што се научило. На пример, понекогаш би сакале да прикажете коефициент чија вредност била од особен интерес. Покрај извлекување заклучоци, во оваа секција би можеле да ги прикажете насоките за натамошни истражувања. Корисно е да се претпостави дека читателот прво би го прочитал заклучокот со цел да одлучи дали да го прочита остатокот од трудот.

5.7 Стилски препораки

Избраниот наслов на вашиот труд треба да го рефлектира проблемот на истражување. Трудот треба да биде отпечатен со проред 1,5 или најмногу 2. Сите равенки треба да започнуваат во нов ред, да бидат центрирани и редоследно нумерирани на следниов начин: (1), (2), (3) и тн. Поголемите графикони или табели можат да бидат прикажани после главниот дел од трудот. Во трудот треба да се реферирате на други трудови според авторот и датумот на пример White (1980). Секцијата на референци на крајот од трудот би требало да биде изработена во стандарден формат.

Кога во секцијата за економетриски модели вклучувате само една равенка, би требало да ги опишете позначајните променливи: зависната променлива и клучната независна променлива или променливи. Со цел да се фокусирате на една единствена независна променлива можете да напишете равенка од следниов тип:

$$Y = \beta_0 + \beta_1 X_1 + \sum_{i=2}^{k} \beta_i X_i + u \qquad \text{или} \qquad Y = \beta_0 + \beta_1 X_1 + \boldsymbol{x\delta} + u$$

При тоа, во равенката одделно фигурира само независната променлива X_1, додека останатите променливи во оваа фаза треба да бидат опишани само генерално. Тие можат подетално да се опишат во посебна табела во секцијата за податоци. На пример, во истражувањето на факторите што влијаат врз висината на платата во секцијата за податоци можете да ја вклучите следнава табела:

Променлива	Опис
Y	Годишна плата (вклучени и додатоци) за 1990 год. (во 1000$)
X_1	Број на години на образование
X_2	Број на години на изминат стаж
X_3	Реализација на претпријатието (во милиони долари)
...	...

Табелата која се однесува на дескриптивните статистики за променливите би го имала следниов изглед:

Променлива	Аритметичка средина	Стандардна девијација	Минимална вредност	Максимална вредност
Y				
X_1				
X_2				
X_3				
X_4				
Број на опсервации:				

Во секцијата на резултатите, оценките можат да бидат претставени било во форма на равенка или во табеларен облик. Табелите особено се корисни кога се оценети повеќе модели со различни групи независни променливи. Доколку оценките ги пишувате во фрома на равенка како на пример:

$$Y = 2{,}45 + 0{,}236X_1 + 0{,}008X_2 + 0{,}061X_3$$
$$(0{,}93)\ (0{,}115)\quad (0{,}003)\quad (0{,}028)$$
$$n = 204 \qquad R^2 = 0{,}351$$

треба да бидете сигурни дека сте назначиле дека во заградите се претставени стандардните грешки. Исто така, за тестирање на нултата хипотеза $H_0 : \beta_i = 0$, прифатливо е да се прикажат и *t*-статистиките или нивните апсолутни вредности со напомена за она што е прикажано.

Доколку резултатите ги прикажувате во табеларна форма, треба јасно да ги назначите зависната и независните променливи. Повторно, треба да се напомене дали под оценетите коефициенти се прикажани стандардните грешки или *t*-статистиките (стандардните грешки имаат предимство). Некои автори користат ѕвездички со цел да укажат на статистичка значајност на различни нивоа на значајност. (на пример, една ѕвездичка означува значајност на ниво од 10%, две ѕвездички означуваат значајност на ниво од 5% и тн.). Тоа не е неопходно да се направи доколку во текстот внимателно ја дискутирате статистичката значајност на независните променливи.

Резултатите би биле лесни за читање и интерпретација доколку единиците мерки на зависната и независните променливи се избрани така што коефициентите не се премногу големи или премногу мали. Доколку оценетите коефициенти се екстремно големи или мали потребно е променливите да бидат изразени во други единици мерки. Исто така, потребно е да се ограничи и бројот на децимални места после децималната запирка. На пример, доколку софтверскиот пакет дава оценки со голем број децимални места, ќе треба тој број да го заокружите на 3 или 2 децимални места.

Општо правило е во трудот да не се прикажуваат командите од софтверскиот пакет што ги користите со цел да ги добиете резултатите. Доколку била користена некоја специфична команда за да се спроведе некој метод за

оценување, тоа може да биде прикажано во прилог. Прилогогот исто така, претставува добар начин за да се вклучат дополнителните резултати што ја поткрепуваат вашата анализа но, не се од суштинско значење.

Табелата со резултатите од оценка на пет различни модели би изгледала на следниов начин:

Независни променливи	1	2	3	4	5
Константа	48.29** (2.33)	62.50** (6.55)	90.94** (6.44)	94.50** (7.32)	117.70** (7.92)
X_1	-5.11* (-1.99)	-6.66** (-3.73)	-3.86* (-1.86)	-4.23** (-2.11)	-4.61** (-2.68)
X_2	4.21 (1.02)				
X_3				-5.00* (-1.72)	
X_4	-0.29 (-0.77)		-0.46 (-1.49)		-0.64** (-2.42)
X_5		-0.42* (-1.95)		-0.46** (-2.37)	-0.79** (-4.15)
X_6		0.12 (0.74)			
X_7			-11.13** (-2.87)		
X_8					-16.43** (-3.76)
X_9				-7.86** (-2.15)	
Прилагоден коеф. на детерминација	0.28	0.38	0.47	0.51	0.64