

Μεταπτυχιακό Πρόγραμμα Σπουδών

Εφαρμοσμένη Οικονομική & Ανάλυση Δεδομένων

Εξεταστική Περίοδος: Χειμερινό Εξάμηνο 2015-2016

Ημερομηνία: 9 Φεβρουαρίου 2016

Μάθημα: Εφαρμοσμένη Μικρο-οικονομετρία

Διδάσκων: Νίκος Γιαννακόπουλος, Επίκουρος Καθηγητής

Ερωτήσεις & Ενδεικτικές Απαντήσεις

Ερώτηση 1. Ποιες είναι οι βασικές ερωτήσεις που πρέπει να απαντήσει ο ερευνητής έτσι ώστε να διασφαλίσει την επιτυχία στην διεκπεραίωση ενός ερευνητικού έργου στις κοινωνικές επιστήμες; (μονάδες 1,0)

Απάντηση 1. Βλέπε: Angrist Joshua D. and Jörn-Steffen Pischke, 2009, Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Ερώτηση 2. Χρησιμοποιώντας ένα παράδειγμα (που εσείς προτιμάτε) παρουσιάστε τις έννοιες «σχέση αιτιώδους συνάφειας» και «όλων των υπολοίπων παραγόντων σταθερών». (μονάδες 1,0)

Απάντηση 2. Βλέπε: Angrist Joshua D. and Jörn-Steffen Pischke, 2009, Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Ερώτηση 3. Χρησιμοποιώντας τη βάση δεδομένων Q3_wagedata2.xls (δεδομένα για άνδρες εργαζόμενους), εκτιμήστε ένα απλό υπόδειγμα μισθών (λογάριθμος μισθού: $\ln wage$) λαμβάνοντας υπόψη τις πληροφορίες που παρέχονται για τα έτη εκπαίδευσης ($educ$), τα έτη εργασιακής εμπειρίας ($exper$), την προϋπηρεσία σε έτη ($tenure$), το εάν ο εργαζόμενος είναι έγγαμος ($married$), το εάν είναι έγχρωμος ($black$) και εάν διαμένει σε νότια πολιτεία ($south$) και σε αστική περιοχή ($urban$).

Υπο-ερώτημα 3i. Με βάση τον εκτιμημένο συντελεστή για το οικογενειακό καθεστώς μπορείτε να απαντήσετε την ακόλουθη ερώτηση; «Εάν συναντήσετε 2 άνδρες, έναν έγγαμο και έναν άγαμο, ποιος αναμένεται να κερδίζει υψηλότερες αμοιβές, όλων των υπολοίπων παραγόντων σταθερών;» (μονάδες 1,0)

Απάντηση 3i. Με βάση τις εκτιμήσεις του εμπειρικού υποδείγματος μισθών (Εικόνα 1), αναμένουμε ότι ένας έγγαμος άνδρας θα λαμβάνει υψηλότερο ωρομίσθιο σε σχέση με έναν μη-έγγαμο επειδή η μεταβλητή για το οικογενειακό καθεστώς σχετίζεται θετικά με το ύψος του ωρομισθίου. Βεβαίως, αυτή η συσχέτιση δεν σημαίνει ότι το υψηλότερο ωρομίσθιο είναι αποτέλεσμα του γάμου.

	Q3i b/se
educ	0.065*** (0.01)
exper	0.014*** (0.00)
tenure	0.012*** (0.00)
married	0.199*** (0.04)
black	-0.188*** (0.04)
south	-0.091*** (0.03)
urban	0.184*** (0.03)
constant	5.395*** (0.11)
R-sqr	0.253
Obs	935

* p<0.05, ** p<0.01, *** p<0.001

Εικόνα 1. Αποτελέσματα εκτιμήσεων (OLS) των προσδιοριστικών παραγόντων του ωρομισθίου (lwage)

Υπο-ερώτημα 3ii. Εάν ο άγαμος παντρευτεί την επόμενη μέρα, μπορείτε να γνωρίζετε πόσο θα αυξηθούν οι αμοιβές του; (μονάδες 1,0)

Απάντηση 3ii. Με βάση το υπόδειγμα που εκτιμήθηκε παραπάνω δεν μπορούμε να γνωρίζουμε το ύψος της αύξησης των αμοιβών των άγαμου ο οποίος παντρεύτηκε την επόμενη μέρα. Αυτή η αδυναμία οφείλεται στο ότι δεν γνωρίζουμε το εάν ο έγγαμος άνδρας λαμβάνει υψηλότερο ωρομίσθιο επειδή ο γάμος των καθιστά πιο παραγωγικό ή εάν αυτό είναι αποτέλεσμα του ότι το οικογενειακό καθεστώς συσχετίζεται με κάποια άλλα μη παρατηρούμενα χαρακτηριστικά τα οποία θα μπορούσαν να εξηγήσουν τις υψηλότερες αμοιβές (π.χ., ικανότητα).

Υπο-ερώτημα 3.iii. Υποθέστε ότι γνωρίζετε πως οι καλύτερα αμειβόμενοι άνδρες είναι αυτοί οι οποίοι έχουν υψηλή πιθανότητα να είναι έγγαμοι (θετική επιλογή έναντι του γάμου). Μπορείτε να υποστηρίξετε ότι με βάση τον εκτιμημένο συντελεστή για το οικογενειακό καθεστώς γνωρίζετε πλήρως το εύρος της επίδρασης του γάμου στις αμοιβές (σε όρους αιτιώδους συνάφειας); (μονάδες 1,0)

Απάντηση 3iii. Η μεροληψία επιλογής είναι θετική γιατί για παράδειγμα, η ικανότητα αυξάνει την πιθανότητα του γάμου αλλά και των αμοιβών. Ως εκ τούτου, ο OLS εκτιμητής (περιλαμβάνει

την πραγματική επίδραση και την μεροληψία στην επιλογή) είναι πιθανό να είναι μεγαλύτερος σε μέγεθος από την «καθαρή» επίδραση του γάμου στο ύψος του ωρομισθίου (σχέση αιτιότητας). Συνεπώς, δεν μπορούμε να υποστηρίξουμε ότι με βάση τον εκτιμημένο συντελεστή για το οικογενειακό καθεστώς γνωρίζουμε πλήρως το εύρος της επίδρασης του γάμου στις αμοιβές.

Υπο-ερώτημα 3iv. Υποθέστε ότι τα ικανότερα άτομα είναι πιο πιθανό να έχουν αυξημένα επίπεδα εκπαίδευσης και υψηλότερες αμοιβές. Εάν δεν λάβετε υπόψη σε μια συνάρτηση αμοιβών (Ελάχιστα Τετράγωνα) μια μεταβλητή για το επίπεδο των ικανοτήτων, ποια από τις υποθέσεις καταστρατηγείται στην εφαρμογή της μεθόδου των ελαχίστων τετραγώνων; Ποια αναμένεται να είναι η επίπτωση της μεροληψίας στον εκτιμητή σας; (μονάδες 1,0)

Απάντηση 3iv. Η παραπάνω παραδοχή καταστρατηγεί την υπόθεση της ανεξαρτησίας των εξηγηματικών μεταβλητών με το διαταρακτικό όρο αφού η εκπαίδευση τώρα συσχετίζεται με το διαταρακτικό όρο. Η διαφορά μεταξύ του εκτιμημένου συντελεστή στο «πραγματικό» υπόδειγμα και στο υπόδειγμα στο οποίο παραλείπεται η μεταβλητή της ατομικής ικανότητας καλείται ως μεροληψία λόγω των παραλειπόμενων μεταβλητών. Επειδή η ικανότητα επηρεάζει θετικά τόσο το ωρομίσθιο όσο και την εκπαίδευση συμπεραίνουμε ότι ο εκτιμητής μας είναι μεροληπτικός και η κατεύθυνση της μεροληψίας αυτής είναι θετική. Αυτό σημαίνει ότι ο OLS εκτιμητής είναι μεγαλύτερος από την πραγματική επίδραση της εκπαίδευσης στο ωρομίσθιο.

Υπο-ερώτημα 3v. Στο παραπάνω υπόδειγμα που εκτιμήσατε εισάγετε την μεταβλητή που προσεγγίζει το επίπεδο ικανοτήτων των εργαζομένων (IQ). Συγκρίνετε τον εκτιμημένο συντελεστή της εκπαίδευσης με αυτόν στο οποίο έχετε παραλείψει τη μεταβλητή IQ. Παρουσιάστε τα αποτελέσματα. Ποια είναι η υπόθεση που θα πρέπει να υιοθετήσετε ώστε να παρουσιάσετε την εκτιμημένη παράμετρο της εκπαίδευσης σε όρους αιτιώδους συνάφειας; Εάν σας πληροφορήσει κάποιος ότι η μέτρηση της μεταβλητής IQ έγινε αφότου το άτομο έχει ολοκληρώσει τις σπουδές του ποια αναμένετε να είναι η ποιότητα αυτής της εκτιμημένης παραμέτρου στην εξήγηση της διακύμανσης των αμοιβών; (μονάδες 1,0)

Απάντηση 3v. Στην Εικόνα (2) παρουσιάζονται τα αποτελέσματα της εκτίμησης του υποδείγματος που παρουσιάζεται στην εκφώνηση του υπο-ερωτήματος 3v. Συγκρίνοντας με τα αποτελέσματα που παρουσιάζονται στην Εικόνα (1) παρατηρούμε ότι η επίδραση της εκπαίδευσης στο ωρομίσθιο μειώνεται από 0.065 σε 0.054 (λόγω της συμπερίληψης της μεταβλητής για το επίπεδο της ατομικής ικανότητας). Αυτό σημαίνει ότι τα άτομα που έχουν υψηλές τιμές στο δείκτη νοημοσύνης (IQ) έχουν υψηλότερη πιθανότητα να βρίσκονται εντός της εκπαίδευσης για περισσότερα χρόνια και να λαμβάνουν επίσης και υψηλότερες αμοιβές. Η μεροληψία λόγω των παραλειπόμενων μεταβλητών στην αρχική εκτίμηση (Εικόνα 1) είναι θετική αφού το IQ επηρεάζει θετικά τα έτη εκπαίδευσης και ταυτόχρονα το IQ συσχετίζεται θετικά και με το ωρομίσθιο.

Για την παρουσίαση της εκτίμησης της παραμέτρου της εκπαίδευσης σε όρους αιτιώδους συνάφειας θα πρέπει να υποθέσουμε ότι λαμβάνοντας υπόψη τις μεταβλητές ελέγχου (exper tenure married black south urban IQ), το επίπεδο του ωρομισθίου των ατόμων με διαφορετικά έτη εκπαίδευσης θα εξηγείται αποκλειστικά από τις διαφορές στα παρατηρούμενα έτη εκπαίδευσης. Με άλλα λόγια, δεν υπάρχουν άλλοι (μη-παρατηρούμενοι) παράγοντες που συσχετίζονται ταυτόχρονα με την εκπαίδευση και το ωρομίσθιο (υπόθεση ανεξαρτησίας).

Εάν το IQ έχει μετρηθεί μετά την ολοκλήρωση της εκπαίδευσης τότε αυτή η μεταβλητή (IQ) αναμένεται να σχετίζεται με την εκπαίδευση και ως εκ τούτου η ποιότητα της ως επεξηγηματική μεταβλητή είναι ιδιαίτερα χαμηλή.

	Q3i b/se
educ	0.054*** (0.01)
exper	0.014*** (0.00)
tenure	0.011*** (0.00)
married	0.200*** (0.04)
black	-0.143*** (0.04)
south	-0.080** (0.03)
urban	0.182*** (0.03)
IQ	0.004*** (0.00)
constant	5.176*** (0.13)
R-sqr	0.263
Obs	935

* p<0.05, ** p<0.01, *** p<0.001

Εικόνα 2. Αποτελέσματα εκτιμήσεων (OLS) των προσδιοριστικών παραγόντων του ωρομισθίου (lwage)

Υπο-ερώτημα 3vi. Στη βάση δεδομένων συμπεριλαμβάνεται μια μεταβλητή για τα έτη εκπαίδευσης του πατέρα του ερωτώμενου (feduc). Πιστεύετε ότι αυτή η μεταβλητή θα μπορούσε να χρησιμοποιηθεί ως μια βοηθητική μεταβλητή για την εκπαίδευση του ερωτώμενου; Εξηγείστε την απάντησή σας. Τι θα πρέπει να υποθέσουμε ώστε η βοηθητική μεταβλητή να είναι έγκυρη; Εάν στο παραπάνω υπόδειγμα συμπεριλάβετε αυτή τη μεταβλητή τι

παρατηρείται για την εκτίμηση της παραμέτρου της εκπαίδευσης του ερωτώμενου (IV-2SLS); (μονάδες 1,0)

Για να είναι αξιόπιστη μια βοηθητική μεταβλητή (z) θα πρέπει αυτή να συσχετίζεται με την ενδογενή μεταβλητή (y), δηλαδή $\text{corr}(y,z) \neq 0$ και ταυτόχρονα θα πρέπει ο διαταρακτικός όρος (ε) να μην σχετίζεται με την ενδογενή μεταβλητή, δηλαδή $\text{corr}(y,\varepsilon) = 0$. Σε ότι αφορά την πρώτη σχέση αυτή θα μπορούσε να ελεγχθεί χρησιμοποιώντας την εκπαίδευση του πατέρα ως βοηθητική μεταβλητή και την εκπαίδευση του ερωτώμενου ως ενδογενή μεταβλητή. Σε ότι αφορά τη δεύτερη σχέση αυτή δεν θα μπορούσε να ελεγχθεί αφού σύμφωνα με αυτή την υπόθεση η εκπαίδευση του πατέρα δεν μπορεί να επηρεάζει τις αμοιβές των παιδιού του παρά μόνον μέσω των ετών εκπαίδευσης του παιδιού. Όμως, αυτός ο περιορισμός δεν είναι ρεαλιστικός αφού οι γονείς με υψηλά επίπεδα εκπαίδευσης θα μπορούσαν να επηρεάζουν τις αμοιβές των παιδιών τους μέσω άλλων δικτύων εκτός της εκπαίδευσης (π.χ., καλύτερη διατροφή κατά την νηπιακή ηλικία, κοινωνικά δίκτυα, μετανάστευση, επιλογή επαγγέλματος κ.λπ.).

Στην Εικόνα 3 παρουσιάζονται τα αποτελέσματα της εκτίμησης του υποδείγματος ωρομισθίου (όπως παρουσιάζεται στην Εικόνα 2) τα οποία έχουν προκύψει από την εφαρμογή του εκτιμητή IV-2SLS. Στην εκτίμηση αυτή η εκπαίδευση του πατέρα είναι η βοηθητική μεταβλητή και η εκπαίδευση του ερωτώμενου είναι η ενδογενής μεταβλητή. Από τα αποτελέσματα του πρώτου σταδίου (1st stage results) παρατηρούμε ότι τα έτη εκπαίδευσης του πατέρα είναι στατιστικά σημαντική μεταβλητή (τουλάχιστον στο 1%) και συσχετίζονται θετικά με τα έτη εκπαίδευσης του ερωτώμενου. Σε ότι αφορά την παλινδρόμηση του ενδιαφέροντος μας ($\ln \text{wage}$) παρατηρούμε ότι η εκπαίδευση του ερωτώμενου είναι στατιστικά σημαντική μεταβλητή (τουλάχιστον στο 1%) και συσχετίζεται θετικά με το ύψος του ωρομισθίου. Συγκρίνοντας την εκτίμηση αυτή με την αντίστοιχη που έχει παρουσιάζεται στην Εικόνα 2 παρατηρούμε ότι τώρα η εκτίμηση της εκπαίδευσης είναι ιδιαίτερα διογκωμένη (0.107 vs. 0.054). Το ίδιο συμβαίνει και με το ύψος του τυπικού σφάλματος (0.031 vs. 0.010). Βεβαίως, δεν μπορούμε να καταλήξουμε σε πλήρες συμπέρασμα για την εξεταζόμενη σχέση αφού η σύγκριση δεν είναι αξιόπιστη λόγω της μείωσης του αριθμού των παρατηρήσεων από 935 (Εικόνα 2) σε 741 (Εικόνα 3).

. ivreg lwage (educ=feduc) exper tenure married black south urban iq, first

First-stage regressions

Source	SS	df	MS	Number of obs = 741		
Model	1669.09703	8	208.637129	F(8, 732) = 74.33		
Residual	2054.517	732	2.80671722	Prob > F = 0.0000		
Total	3723.61404	740	5.03191086	R-squared = 0.4482		
				Adj R-squared = 0.4422		
				Root MSE = 1.6753		

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	-.1637717	.0158341	-10.34	0.000	-.1948574	-.132686
tenure	.0205821	.0128137	1.61	0.109	-.004574	.0457382
married	-.0731816	.1964687	-0.37	0.710	-.4588908	.3125277
black	.3165394	.2327815	1.36	0.174	-.1404596	.7735383
south	.1136194	.1370728	0.83	0.407	-.1554834	.3827222
urban	.2195851	.1392515	1.58	0.115	-.0537948	.4929651
iq	.0624148	.0047636	13.10	0.000	.053063	.0717667
feduc	.1460854	.0206711	7.07	0.000	.1055037	.1866672
_cons	7.311459	.5926226	12.34	0.000	6.148016	8.474902

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs = 741		
Model	25.1930715	8	3.14913393	F(8, 732) = 24.74		
Residual	103.857102	732	.141881287	Prob > F = 0.0000		
Total	129.050173	740	.174392126	R-squared = 0.1952		
				Adj R-squared = 0.1864		
				Root MSE = .37667		

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1078224	.0318141	3.39	0.001	.0453646	.1702802
exper	.0268477	.006855	3.92	0.000	.01339	.0403055
tenure	.0074962	.0029405	2.55	0.011	.0017235	.0132689
married	.1940357	.0442384	4.39	0.000	.1071865	.2808849
black	-.1127266	.0522905	-2.16	0.031	-.215384	-.0100693
south	-.077322	.030601	-2.53	0.012	-.1373982	-.0172459
urban	.1627543	.0327586	4.97	0.000	.0984423	.2270664
iq	.0004337	.0024719	0.18	0.861	-.0044192	.0052866
_cons	4.673668	.2921805	16.00	0.000	4.100056	5.24728

Instrumented: educ

Instruments: exper tenure married black south urban iq feduc

Εικόνα 3. Αποτελέσματα εκτιμήσεων (IV-2SLS) των προσδιοριστικών παραγόντων του ωρομισθίου (lwage)

Ερώτηση 4. Χρησιμοποιώντας τη βάση δεδομένων Q4_wagepanel.xls (δεδομένα διαστρωματικών χρονοσειρών για εργαζόμενους άνδρες), εκτιμήστε ένα απλό υπόδειγμα μισθών ($lwage$) λαμβάνοντας υπόψη τις πληροφορίες που παρέχονται για τα έτη εκπαίδευσης ($educ$), τα έτη εργασιακής εμπειρίας ($exper$) σε μορφή πολυωνύμου δευτέρου βαθμού, το εάν ο εργαζόμενος είναι έγγαμος ($married$), το εάν είναι έγχρωμος ($black$), εάν είναι Ισπανόφωνος ($hisp$) εάν είναι μέλος εργατικού σωματείου ($union$) και για τα έτη διενέργειας της έρευνας (χρονικές επιδράσεις). Εκτιμήστε ένα υπόδειγμα το οποίο αγνοεί τη χρονική διάσταση των δεδομένων (Ελάχιστα Τετράγωνα) καθώς και ένα υπόδειγμα σταθερών επιδράσεων. Συγκρίνετε τις εκτιμημένες παραμέτρους της συμμετοχής στο εργατικό σωματείο μεταξύ των δυο υποδειγμάτων. Γιατί πιστεύετε ότι υπάρχουν διαφορές στους εκτιμημένους συντελεστές; Συμβαίνει το ίδιο και με τη μεταβλητή για το εάν ο εργαζόμενος είναι έγγαμος; Εάν ναι που μπορεί να οφείλονται οι διαφορές; (μονάδες 1,0)

Απάντηση 4. Στην Εικόνα 4α παρουσιάζονται τα αποτελέσματα των εκτιμήσεων που αφορούν το υπόδειγμα που αγνοεί τη χρονική διάσταση των δεδομένων (OLS). Στην Εικόνα 4β παρουσιάζονται τα αποτελέσματα των εκτιμήσεων που αφορούν το υπόδειγμα που λαμβάνει υπόψη τη χρονική διάσταση των δεδομένων (υπόδειγμα σταθερών επιδράσεων, FE). Σε ότι αφορά το αρχικό υπόδειγμα (OLS) παρατηρούμε ότι οι εργαζόμενοι που είναι μέλη εργατικού σωματείου λαμβάνουν, κατά μέσο όρο, 18% υψηλότερες αμοιβές από τους συναδέλφους τους που δεν είναι μέλη εργατικού σωματείου (με κατά τα άλλα παρόμοια χαρακτηριστικά όπως τα, $race$, $work\ experience$, $marital\ status$ and $education$). Στο δεύτερο υπόδειγμα (FE) παρατηρούμε ότι η μεταβλητή που καταδεικνύει τη συμμετοχή σε εργατικό σωματείο είναι στατιστικά σημαντική (τουλάχιστον στο 1%) αλλά το ύψος της εκτίμησης είναι σαφώς μικρότερο (8%) από το αντίστοιχο στην Εικόνα 4α. Η υπερεκτίμηση της αρχικής παλινδρόμησης (OLS) οφείλεται στην αδυναμία του εμπειρικού υποδείγματος να λάβει υπόψη το γεγονός ότι κάποιοι εργαζόμενοι με συγκεκριμένα χαρακτηριστικά (π.χ., με υψηλά επίπεδα ικανοτήτων και δυνατότητα αναγνώρισης κινήτρων) έχουν επιλέξει στο παρελθόν να είναι μέλη του εργατικού σωματείου και ως εκ τούτου να λαμβάνουν διαχρονικά υψηλές αμοιβές (λόγω αυτής της επιλογής). Το υπόδειγμα FE που αναγνωρίζει το γεγονός ότι τα ίδια άτομα παρατηρούνται στο χρόνο αντιμετωπίζει το πρόβλημα της χρονικά αμετάβλητης μη-παρατηρούμενης ετερογένειας και εξαλείφει την υπερεκτίμηση μέσω της χρήσης σταθερών επιδράσεων σε επίπεδο εργαζόμενου. Έτσι λαμβάνονται υπόψη οι χρονικά σταθερές διαφορές μεταξύ όσων εργαζόμενων είναι μέλη εργατικού σωματείου και όσων δεν είναι.

Σε ότι αφορά τη σύγκριση της εκτίμησης της μεταβλητής της οικογενειακής κατάστασης μεταξύ των δύο υποδειγμάτων παρατηρούμε ότι στην πρώτη περίπτωση η εκτίμηση ανέρχεται σε 0.108 (Εικόνα 4α) ενώ στη δεύτερη περίπτωση υποχωρεί σε 0.046 (Εικόνα 4β). Και στις δυο περιπτώσεις η μεταβλητή $married$ είναι στατιστικά σημαντική (τουλάχιστον στο 1% στην OLS και τουλάχιστον στο 5% στην FE). Οι διαφορές οφείλονται στο γεγονός ότι στην περίπτωση των ομαδοποιημένων στοιχείων χωρίς χρονική διάσταση το μισθολογικό πλεονέκτημα των εγγάμων αντικατοπτρίζει μηχανισμούς θετικής επιλογής (οι πιο ικανοί έχουν μεγαλύτερη πιθανότητα να είναι έγγαμοι).

```
. reg lwage black exper exp2 hisp married educ union d82 d83 d84 d85 d86 d87
```

Source	SS	df	MS			
Model	233.196533	13	17.9381948	Number of obs = 4360		
Residual	1003.33309	4346	.230863573	F(13, 4346) = 77.70		
				Prob > F = 0.0000		
				R-squared = 0.1886		
				Adj R-squared = 0.1862		
Total	1236.52962	4359	.283672774	Root MSE = .48048		

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
black	-.1398484	.0235847	-5.93	0.000	-.1860864	-.0936104
exper	.0743651	.0131865	5.64	0.000	.0485129	.1002173
exp2	-.002759	.0008	-3.45	0.001	-.0043274	-.0011905
hisp	.0162835	.0208031	0.78	0.434	-.0245013	.0570682
married	.108592	.0156933	6.92	0.000	.0778251	.1393588
educ	.0921175	.0052237	17.63	0.000	.0818764	.1023587
union	.1816199	.0171565	10.59	0.000	.1479844	.2152554
d82	.0272389	.0275967	0.99	0.324	-.0268647	.0813425
d83	.0231358	.0305801	0.76	0.449	-.0368168	.0830885
d84	.0489718	.0337867	1.45	0.147	-.0172674	.115211
d85	.0658025	.0370003	1.78	0.075	-.0067371	.138342
d86	.0972761	.04019	2.42	0.016	.0184831	.176069
d87	.1286576	.0434966	2.96	0.003	.043382	.2139332
_cons	.092606	.0782938	1.18	0.237	-.0608898	.2461018

Εικόνα 4α. Αποτελέσματα εκτιμήσεων (OLS) των προσδιοριστικών παραγόντων του ωρομισθίου (lwage)

```
. xtreg lwage black exper exp2 hisp married educ union d82 d83 d84 d85 d86 d87, fe
note: black omitted because of collinearity
note: hisp omitted because of collinearity
note: educ omitted because of collinearity
```

```
Fixed-effects (within) regression                Number of obs   =    4360
Group variable: nr                             Number of groups =    545

R-sq:  within = 0.1806                        Obs per group:  min =     8
        between = 0.0001                       avg   =    8.0
        overall = 0.0481                       max   =     8

                                                F(10, 3805)    =    83.85
corr(u_i, Xb) = -0.2018                       Prob > F       =    0.0000
```

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
black	0	(omitted)				
exper	.1511912	.0219489	6.89	0.000	.1081584	.194224
exp2	-.0051855	.0007044	-7.36	0.000	-.0065666	-.0038044
hisp	0	(omitted)				
married	.0466804	.0183104	2.55	0.011	.0107812	.0825796
educ	0	(omitted)				
union	.0800019	.0193103	4.14	0.000	.0421423	.1178614
d82	-.0494115	.0368581	-1.34	0.180	-.121675	.0228519
d83	-.0991298	.0564143	-1.76	0.079	-.209735	.0114753
d84	-.11465	.077131	-1.49	0.137	-.2658721	.0365721
d85	-.1384738	.0984555	-1.41	0.160	-.3315044	.0545569
d86	-.1416506	.120262	-1.18	0.239	-.3774348	.0941335
d87	-.1333135	.1425382	-0.94	0.350	-.4127721	.1461452
_cons	.9702256	.075978	12.77	0.000	.821264	1.119187
sigma_u	.41191781					
sigma_e	.35099001					
rho	.57935646	(fraction of variance due to u_i)				

```
F test that all u_i=0:      F(544, 3805) =    9.21      Prob > F = 0.0000
```

Εικόνα 4β. Αποτελέσματα εκτιμήσεων (FE) των προσδιοριστικών παραγόντων του ωρομισθίου (lwage)

Ερώτηση 5. Χρησιμοποιώντας τη βάση δεδομένων Q5_binlfr2.xls (δεδομένα έγγαμων γυναικών), εκτιμήστε ένα απλό υπόδειγμα συμμετοχής (probit) στο εργατικό δυναμικό (lfr=1) λαμβάνοντας υπόψη τις πληροφορίες που παρέχονται για την ηλικία της συζύγου σε έτη (age), το εάν έχει ολοκληρώσει την τριτοβάθμια εκπαίδευση (wc=1), τον αριθμό των παιδιών ηλικίας 0-5 (k5), τον αριθμό των παιδιών ηλικίας 6-15 (k618), το εάν έχει ο σύζυγος έχει ολοκληρώσει την τριτοβάθμια εκπαίδευση (wc=1), το λογάριθμο του μισθού του συζύγου (lwg) και το εισόδημα της οικογένειας εξαιρουμένου του μισθού του συζύγου (inc). Παρουσιάστε αναλυτικά τα αποτελέσματα των εκτιμήσεων σας (στατιστική και οικονομική ερμηνεία). Λόγω της μη-γραμμικής επίδρασης που ασκεί η ηλικία στην πιθανότητα συμμετοχής των γυναικών στην αγορά εργασίας χρησιμοποιείστε ψευδομεταβλητές για 5ετείς ηλικιακές ομάδες (αντί της μεταβλητής age) και επανεκτιμήστε το παραπάνω υπόδειγμα χρησιμοποιώντας το δείγμα των έγγαμων γυναικών ηλικίας 30-54 ετών. Παρουσιάστε αναλυτικά τα αποτελέσματα των

εκτιμήσεων σας (στατιστική και οικονομική ερμηνεία) και συγκρίνετε με τα αποτελέσματα του αρχικού υποδείγματος. (μονάδες 1,0)

Απάντηση 5. Στην Εικόνα 5α παρουσιάζονται οι εκτιμημένοι συντελεστές των προσδιοριστικών παραγόντων της συμμετοχής στο εργατικό δυναμικό υιοθετώντας ένα υπόδειγμα Probit. Σε όρους στατιστικής παρουσίασης των αποτελεσμάτων παρατηρούμε ότι η εν λόγω εξειδίκευση είναι χρήσιμη για την διερεύνηση της πιθανότητας συμμετοχής στο εργατικό δυναμικό (LR-test) και μάλιστα το επίπεδο σημαντικότητας των στατιστικά σημαντικών μεταβλητών ($k618$ και hc) είναι τουλάχιστον το 1%. Ειδικότερα, παρατηρούμε ότι όσο μεγαλύτερη είναι η ηλικία της έγγαμης γυναίκας τόσο μικρότερη είναι η πιθανότητα ένταξης της στην αγορά εργασίας. Οι έγγαμες γυναίκες με τριτοβάθμια εκπαίδευση έχουν αυξημένη πιθανότητα συμμετοχής στο εργατικό δυναμικό ενώ αντίθετα όσο μεγαλύτερος είναι ο αριθμός των παιδιών ηλικίας 0-5 ετών τόσο μικρότερη η συμμετοχή στην αγορά εργασίας. Επίσης, παρατηρούμε ότι οι γυναίκες των οποίων ο σύζυγος έχει αυξημένες απολαβές από εργασία εμφανίζουν αυξημένη πιθανότητα ένταξης στην αγορά εργασίας. Αντίθετα, η πιθανότητα μια έγγαμη γυναίκα να εισέλθει στην αγορά εργασίας είναι μειωμένη όσο υψηλότερο είναι το εισόδημα του νοικοκυριού από πηγές εκτός εργασίας.

Σε όρους οικονομικής ερμηνείας των εκτιμήσεων παρατηρούμε από τον πίνακα των οριακών επιδράσεων (Εικόνα 4β) ότι η πιθανότητα συμμετοχής στο εργατικό δυναμικό μειώνεται κατά 1.4% για ένα επιπρόσθετο έτος στην ηλικία, κατά 34.2% για ένα επιπρόσθετο παιδί ηλικίας 0-5 ετών και κατά 0.8% εάν το οικογενειακό εισόδημα αυξηθεί κατά 1.000 χρηματικές μονάδες ετησίως. Αντίθετα, η πιθανότητα συμμετοχής στο εργατικό δυναμικό είναι μεγαλύτερη κατά 18 ποσοστιαίες μονάδες για όσες γυναίκες έχουν ολοκληρώσει τριτοβάθμια εκπαίδευση (σε σχέση με όσες δεν έχουν ολοκληρώσει τριτοβάθμια εκπαίδευση). Επίσης, εάν το εισόδημα από εργασία του συζύγου αυξηθεί κατά 1.000 χρηματικές μονάδες ετησίως η πιθανότητα εισόδου της συζύγου στην αγορά εργασίας αυξάνεται κατά 14.3%.

Στην Εικόνα 5γ παρουσιάζονται οι εκτιμήσεις του υποδείγματος της συμμετοχής στο εργατικό δυναμικό για έγγαμες γυναίκες ηλικίας 30-54 ετών εισάγοντας ψευδομεταβλητές για την ηλικία της καθεμιάς γυναίκας (πενταετείς ομάδες ηλικίας). Θα πρέπει να σημειώσουμε ότι τώρα το δείγμα είναι μικρότερο σε σχέση με αυτό που χρησιμοποιήθηκε για τις εκτιμήσεις που παρουσιάστηκαν στην Εικόνα 5α (698 vs. 753). Αυτό οφείλεται στο ότι το αρχικό δείγμα αναφέρεται σε γυναίκες ηλικίας 30-60 ετών ενώ τώρα το δείγμα περιορίζεται σε αυτές στην ηλικιακή ομάδα των 30-54 ετών. Θα πρέπει να σημειώσουμε ότι σε στατιστικούς όρους δεν παρατηρούμε σημαντικές διαφοροποιήσεις μεταξύ των εκτιμημένων αποτελεσμάτων με τη μόνη διαφορά στην περίπτωση της ηλικίας. Όταν χρησιμοποιούμε ψευδομεταβλητές παρατηρούμε ότι η επίπτωση της ηλικίας στην πιθανότητα συμμετοχής δεν είναι γραμμική αλλά μη-γραμμική και μάλιστα οι διαφορές εμφανίζονται ως σημαντικές μετά την ηλικία των 40 ετών (σε σχέση με την ηλικιακή ομάδα των 30-34 ετών). Σε οικονομικούς όρους (Εικόνα 5δ, οριακές επιδράσεις) η πιθανότητα συμμετοχής για τις γυναίκες ηλικίας 40-44 ετών σε σχέση με αυτές στην ομάδα των 30-34 ετών είναι μειωμένη κατά 18 ποσοστιαίες μονάδες. Αντιστοίχως, για όσες είναι στην ηλικιακή ομάδα των 45-49 ετών η πιθανότητα είναι μειωμένη κατά 16 ποσοστιαίες μονάδες ενώ και για όσες είναι στην ηλικιακή ομάδα των 50-54 ετών η πιθανότητα είναι μειωμένη κατά 29 ποσοστιαίες μονάδες.

```

. probit lfp age wc k5 k618 hc lwg inc

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -452.84838
Iteration 2:   log likelihood = -452.69498
Iteration 3:   log likelihood = -452.69496

Probit regression                               Number of obs   =       753
                                                LR chi2(7)      =       124.36
                                                Prob > chi2     =       0.0000
Log likelihood = -452.69496                    Pseudo R2      =       0.1208

```

lfp	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.0378235	.0076093	-4.97	0.000	-.0527375 -.0229095
wc	.4883144	.1354873	3.60	0.000	.2227641 .7538647
k5	-.8747111	.1135584	-7.70	0.000	-1.097281 -.6521408
k618	-.0385945	.0404893	-0.95	0.340	-.1179521 .0407631
hc	.0571703	.1240053	0.46	0.645	-.1858755 .3002162
lwg	.3656287	.0877792	4.17	0.000	.1935846 .5376727
inc	-.020525	.0047769	-4.30	0.000	-.0298875 -.0111625
_cons	1.918422	.3806539	5.04	0.000	1.172354 2.66449

Εικόνα 5α. Αποτελέσματα εκτιμήσεων (Probit) των προσδιοριστικών παραγόντων της συμμετοχής στο εργατικό δυναμικό (lfp)

```

. mfx compute

Marginal effects after probit
y = Pr(lfp) (predict)
= .57816367

variable      dy/dx      Std. Err.      z      P>|z|      [ 95% C.I. ]      X
-----
age           -.0147988     .00297        -4.98     0.000     -.020626 -.008971     42.5378
wc*           .1843553     .04856         3.80     0.000     .089173 .279538     .281541
k5            -.3422398     .04453        -7.69     0.000     -.429507 -.254972     .237716
k618         -.0151005     .01584         -0.95     0.340     -.046147 .015946     1.35325
hc*          .0223379     .04838         0.46     0.644     -.072492 .117168     .391766
lwg           .143056      .03426         4.18     0.000     .075908 .210204     1.09711
inc          -.0080306     .00187         -4.30     0.000     -.011695 -.004366     20.129

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```

Εικόνα 5β. Οριακές επιδράσεις του εκτιμημένου υποδείγματος (Probit) των προσδιοριστικών παραγόντων της συμμετοχής στο εργατικό δυναμικό (lfp)

```
. probit lfp ac_2 ac_3 ac_4 ac_5 wc k5 k618 hc lwg inc if age>=30 & age<=54
```

```
Iteration 0: log likelihood = -475.11287
Iteration 1: log likelihood = -419.76821
Iteration 2: log likelihood = -419.62573
Iteration 3: log likelihood = -419.62572
```

```
Probit regression                               Number of obs   =       698
                                                LR chi2(10)    =      110.97
                                                Prob > chi2    =       0.0000
Log likelihood = -419.62572                    Pseudo R2      =       0.1168
```

lfp	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ac_2	-.0983297	.1671134	-0.59	0.556	-.4258659	.2292065
ac_3	-.4695712	.1724141	-2.72	0.006	-.8074966	-.1316458
ac_4	-.4132676	.1702808	-2.43	0.015	-.747012	-.0795233
ac_5	-.7495148	.187789	-3.99	0.000	-1.117574	-.3814551
wc	.4992049	.141545	3.53	0.000	.2217819	.776628
k5	-.8454866	.1146886	-7.37	0.000	-1.070272	-.620701
k618	-.0397372	.0429466	-0.93	0.355	-.1239109	.0444365
hc	-.0043276	.1281674	-0.03	0.973	-.255531	.2468759
lwg	.3454296	.0919917	3.76	0.000	.1651292	.52573
inc	-.018982	.0049696	-3.82	0.000	-.0287222	-.0092418
_cons	.6848747	.1872129	3.66	0.000	.3179441	1.051805

Εικόνα 5γ. Αποτελέσματα εκτιμήσεων (Probit) των προσδιοριστικών παραγόντων της συμμετοχής στο εργατικό δυναμικό (lfp), ηλικίες 30-54

```
. mfx compute
```

```
Marginal effects after probit
```

```
y = Pr(lfp) (predict)
= .58860994
```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]		X
ac_2*	-.0384819	.06573	-0.59	0.558	-.167308	.090344	.194842
ac_3*	-.1850997	.06746	-2.74	0.006	-.317319	-.05288	.184814
ac_4*	-.1626832	.0669	-2.43	0.015	-.293799	-.031567	.230659
ac_5*	-.2920104	.0693	-4.21	0.000	-.427834	-.156187	.157593
wc*	.1867067	.05004	3.73	0.000	.088622	.284791	.282235
k5	-.3289456	.04475	-7.35	0.000	-.416644	-.241247	.256447
k618	-.0154602	.01671	-0.93	0.355	-.048207	.017287	1.45415
hc*	-.0016839	.04987	-0.03	0.973	-.099435	.096068	.398281
lwg	.1343931	.03569	3.77	0.000	.064433	.204353	1.09532
inc	-.0073852	.00193	-3.82	0.000	-.011176	-.003594	20.2529

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Εικόνα 5δ. Οριακές επιδράσεις του εκτιμημένου υποδείγματος (Probit) των προσδιοριστικών παραγόντων της συμμετοχής στο εργατικό δυναμικό (lfp), ηλικίες 30-54

Παράρτημα εντολών Stata

**Use the Q3_wagedata2 file

*Question 3i

reg lwage educ exper tenure married black south urban

estimates store m1, title(Q3i)

estout m1, cells(b(star fmt(3)) se(par fmt(2))) ///

legend label varlabels(_cons constant) ///

stats(r2 N, fmt(3 0 1) label(R-sqr Obs))

*Question 3v

reg lwage educ exper tenure married black south urban iq

estimates store m1, title(Q3i)

estout m1, cells(b(star fmt(3)) se(par fmt(2))) ///

legend label varlabels(_cons constant) ///

stats(r2 N, fmt(3 0 1) label(R-sqr Obs))

*Question 3vi

ivreg lwage (educ=feduc) exper tenure married black south urban iq, first

**Use the Q4_wagepanel file

*Question 4

xtset nr year

gen exp2=exper^2

reg lwage black exper exp2 hisp married educ union d82 d83 d84 d85 d86 d87

xtreg lwage black exper exp2 hisp married educ union d82 d83 d84 d85 d86 d87, fe

**Use the Q5_binlfp2

*Question 5

probit lfp age wc k5 k618 hc lwg inc

mfx compute

```
gen agecat=.
replace agecat=1 if age>=30 & age<=34
replace agecat=2 if age>=35 & age<=39
replace agecat=3 if age>=40 & age<=44
replace agecat=4 if age>=45 & age<=49
replace agecat=5 if age>=50 & age<=54
tab agecat, gen(ac_)
probit lfp ac_2 ac_3 ac_4 ac_5 wc k5 k618 hc lwg inc if age>=30 & age<=54
mfx compute
*****
```