Exam SØK3001, spring 2024. Assessment guidelines.

This is guidelines for assessment. Thus, it is not a complete suggestion of solution. The presentation here is shorter than expected for a complete solution.

Question 1 (20%).

Briefly explain the following concepts

- **a) Serially correlated error term**
- **b) Heteroscedastic error term**
- **c) Stationary time series**
- **d) Measurement error**
- **e) Structural form equation**

a) In a time series regression equation $y_t = \beta_0 + \beta_1 x_t + u_t$, serially correlated errors imply that we relax the assumption that the error term in one period (t) is independent of the error term in another periods (s). That is, we assume $cov(u_t, u_s|x) \neq 0,\ \ s \neq t$

b) In a regression model $y_i = \beta_0 + \beta_1 x_i + u_i$, heterocedastitic error term implies that the variance of the error term is not constant. $Var(u_i|x_i) = \sigma_i^2$, i.e., the variation differs across cross section units *i*.

c) Intuition: If stationary, the effect of a shock to a time series y_t will be eliminated, i.e., y_t returns to the mean level after a shock

Formally: A time series y_t is said to be weakly stationary if

- expected value $E(y_t)$ is constant and independent of t: $E(y_t) = \mu,\ \ t = 1,2, \dots \dots, \infty$
- variance $var(y_t)$ is constant and independent of t: $var(y_t) = \sigma^2 = \gamma_0,\ \ t = 1, 2, \dots \dots, \infty$
- covariance only depends on the difference in time, not on time itself $covar(y_t, y_{t+s}) = \gamma_s = covar(y_t, y_{t+s+m})$

d) When we use an imprecise measure of an economic variable in a regression model, then the model contains measurement error. Measurement error may appear in the dependent varible or in one or more of the explanatory variables. Mismeasured explanatory variables are of primary interest, as it may lead to correlation between the error term in the regression model and the mismeasured explanatory variable and thus biased OLS estimators (see ch. 9 in textbook).

e) A structural form equation can be defined as a behavioral equation derived from economic theory (e.g., a demand equation as in Question 2) and the parameters in the equation has a causal interpretation, for example the price elasticity of demand.

Question 2 (40%).

A researcher has estimated the demand function for airline seats in the US based on data for 1149 different routes from the year 1997. A simple demand function for airline seats is

(1) $\mathbf{\log(\boldsymbol{passen}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \log(\boldsymbol{fare}) + \boldsymbol{\beta}_2 \log(\boldsymbol{dist}) + \boldsymbol{\beta}_1 [\log(\boldsymbol{dist})]^2 + \boldsymbol{u}_1}$

where

$passen$= average number of passengers per day

$fare$ =average airfare (price) in US$

$dist$=route distance in miles

a) If (1) is truly a demand function, what should be the sign of β_1?

If (1) is truly a demand function (structural equation), β_1<0 as it is interpreted as the (constant) price elasticity of demand for airline seats. Notice that there is a misprint in equation (1) in that the parameter in front of $\log(dist)^2$ should be β_3.

Table 1 shows results from different estimated equations based on the data set for these variables from the year 1997. lpassen is $log(passen)$, lfare is $log(fare)$, ldist is $log(dist)$, ldistsq is $[\log(dist)]^2$. *vhat* is the residual from the regression in column (3). The variable $concen$ is a measure of market concentration on the route defined as the share of flights conducted by the largest carrier (company). Table 2 presents descriptive statistics for passen, fare, dist and concen.

b) What is the estimated price elasticity of demand according to the results in Table 1?

The estimated price elasticity in the OLS regression in column (1) is -0.391, while the estimated price elasticity in the IV/2SLS regression is -1.174

c) What is the interpretation of the regression equations in column (2) and (3)?

Column (2) is the estimated structural equation (the demand equation for airline seats), while column (3) is the estimated first stage or reduced form equation for the price (log(fare)).

d) Explain how the equation in column (2) is estimated.

The equation in column (2) is estimated by the two-stage least squares method. In the first stage a price equation with lfare as the dependent variable is estimated by OLS, containing the exogenous variables: log distance and log distance squared, in addition to the instrumental variable concen that is assumed to affect prices but not demand directly. In the second stage the structural equation is estimated by OLS with lfare replaced by its predicted value from the first stage (reduced form) equation in column (3).

e) Briefly explain the economic argument for using the variable *concen* as an instrumental variable.

The economic argument is that prices depend on the competion between airline companys on the route, higher competition means lower prices, ceteris parabus. If all flights are done by one company, concen is 1. With more than one company, the value of concen is below 1. The higher the concen variable, the lower is the competition in the market, and the higher is the expected price (lfare). This is also the finding in column (3).

f) Explain what is meant by weak instruments. Use the results in Table 1 to decide whether you have a weak instrument problem in this case

A weak instrument is a situation where the instrumental variable(s) is (are) only weakly correlated with the endogeneous right hand side variable (here: lfare). If this is the case, the IV/2SLS estimator may be no better than the OLS estimator. A necessary condition for the IV/2SLS method to produce credible estimates of the causal effect of prices on airline seats is that correlation between the price (lfare) and the the instrumental variable (concen) is sufficiently high. Whether you have a weak instrumental problem can be detected by a test of the null hypothesis that the coefficients in front of the instruments in the first stage equation are zero. If the null hypothesis is rejected by a clear margin (rule of thumb is an F-statistic>10) it indicates that the instruments are not weak. In our case with one instrument (concen) this implies a t-statistic above 3.2 for the coefficient in front of concen in col (3). The actual t-statistic is 0.395/0.063=6.26 which indicates that there does not seem to be a weak instrument problem here.

g) Explain how you can test whether the price variable, lfare, is exogeneous. Use the results in Table 1 to test the hypothesis that lfare is exogeneous.

Intuitively, if the price variable, lfare is exogeneous, the IV and OLS estimates of lfare would be quite similar as both would be consistent (given that concen is a valid instrument). The test is explained in 15-5a in the textbook. The test is based on an OLS regression of the structural equation, i.e. the demand equation, extended by the estimated residuals from the first stage regression. This extended model is estimated in col (4) in Table 1. Under the null hypothesis of exogeneity of lfare, the coefficient in front of the first stage residual, vhat, should be zero. The t-statistic is 0.810/0.373=2.17 and based on this evidence we reject the hypothesis that lfare is exogeneous at 5% level (critical value is 1.96, see statistical table G.2).

h) A commentator argues that one should test for overidentification restrictions when using the instrumental variable approach. Do you agree with the commentator? Explain your answer.

Testing for overidentification restrictions is only relevant when the structural equation is overidentified. In our case, the structural equation is just (exactly) identified, as we have one endogeneous explanatory variable (lfare) and one instrumental variable (concen). Thus, it is not meaningful to test for overidentification restrictions.

i) Using the results in column (2), describe how demand for seats depends on route distance

The candidates should realize that the formulation in equation (1) implies that the relationship between demand for seats and route distance is nonlinear as expressed by the elasticity of demand with respect to route distance.

Differentiating the demand equation (1) with respect to distance gives the elasticity of demand with respect to distance as (Notice that there is a misprint in equation (1) in that the parameter in front of $\log(\text{dist})^2$ should be β_3).

$$\frac{dpassen}{ddistance}\frac{distance}{passen} = \frac{dlog(passen)}{dlog(distance)} = \beta_2 + 2\beta_3 \log(distance)$$

The equation shows that the elasticity is not constant, but varies with log (distance).

Using the numbers in Table 2, we can compute the estimated elasticity evaluated for different values of distance. Using the estimates in column (2) in Table 1 and the mean value in Table 2, we have that the elasticity is

$$-2.176 + 2 \cdot 0.187 \cdot \log(989.745) \approx 0.40$$

Evaluated at the maximum distance in the sample, the estimated elasticity is

$$-2.176 + 2 \cdot 0.187 \cdot log(2724) \approx 0.70,$$

Evaluated at the minimum distance in the sample, the estimated elasticity is

$$-2.176 + 2 \cdot 0.187 \cdot log(95) \approx -0.47$$

Table 1. Estimation results. Estimated standard errors in parentheses. The text under the column number shows the dependent variable used in the regression.

	(1) lpassen	(2) lpassen	(3) Lfare	(4) lpassen
lfare	-0.391	-1.174		-1.174
	(0.067)	(0.388)		(0.367)
ldist	-1.570	-2.176	-0.936	-2.176
	(0.629)	(0.726)	(0.272)	(0.687)
ldistsq	0.116	0.187	0.108	0.187
	(0.048)	(0.061)	(0.021)	(0.058)
concen			0.395	
			(0.063)	
vhat				0.810
				(0.373)
Constant	13.230	18.014	6.190	18.014
	(2.100)	(3.217)	(0.890)	(3.042)
Instruments	-	*concen*	-	-
Observations	1,149	1,149	1,149	1,149
Method	OLS	IV	OLS	OLS
R-squared	0.057		0.408	0.061

Table 2. Descriptive statistics. Means, standard deviations, min and max

	Observations	Mean	Standard deviation	Minimum	Maximum
passen	1149	601.042	763.5326	27	7637
fare	1149	173.752	76.30483	37	460
dist	1149	989.745	612.0313	95	2724
concen	1149	0.61254	0.198131	0.192	1

Question 3 (40%)

Politicians are concerned of lack of teachers. There are teacher shortages in several municipalities. The municipalities have to employ teachers who are not formally qualified to be teachers. The teacher union argues that one reason for teacher shortages is high job pressure related to many students in the classroom. In order to reduce teacher shortages, the union has argued that the municipalities should employ more teachers in order to reduce the job pressure.

You are asked to investigate the claim of the union. You are given access to data for municipalities where

- **Short = the percent of teachers not formally qualified**
- **logTeacher = the logarithm of the number of teachers**
- **logStudents = the logarithm of the number of students**
- **Central = an index for the centrality of the municipality. The index has the lowest values for rural areas in the periphery and the highest values for the big cities.**

You plan to estimate the following model for the situation in the fall of 2019.

(1) $$\mathbf{Short_i = \beta_0 + \beta_1 \log Teachers_i + \beta_2 Central_i + u_i}$$

where subscript *i* denotes municipality and *u* is the error term.

a) What are the necessary assumptions to obtain unbiased estimators by the Ordinary Least Square (OLS) method?

The assumptions for unbiased coefficients are (i) linearity in parameters, (ii) random sampling, (iii) no perfect collinearity, and (iv) zero conditional mean. In reality, the last assumption is the most challenging in economic analyses. It is expected that some explanations are provided for the assumptions, in particular (ii) and (iv).

b) Formulate hypotheses for the coefficients in the model (1).

It is possible to follow the union's argument, saying that more teachers reduce shortages, i.e., $\beta_1 < 0$. It is also possible to rely on economic theory: Higher demand in a situation with excess demand increases shortages, i.e., $\beta_1 > 0$.

c) The results for the model are presented in column (1) in Table 3. Explain the statistics R-squared (coefficient of determination) and R-squared adjusted.

R-squared is the ratio of the explained variation to the total variation. It is expected that the definitions of the explained variation and the total variation are provided. R-squared adjusted takes the degrees of freedom into account and adds a penalty when adding more variables. Thus, the R-squared adjusted is smaller than the regular R-squared. It might be useful to include the formal definitions in order to be precise.

d) Table 3 presents the estimated coefficients, with standard errors in parentheses. Interpret the findings in column (1). Are they in accordance with your hypotheses?

It is expected that both the parameters for *log(Teachers)* and *Central* are commented. The coefficient for log(Teachers) implies that when the number of teachers increases by 1 log-point, shortages increase with 0.37 percentage points. Or more realistically, increased number of teachears by 10% (which is close to 0.1 log-points), increases shortages by 0.037 percentage points. This is a very low effect. The t-value is 0.37/0.46 = 0.80, which is below the critical value at all reasonable significance levels. The effect *Central* cannot be interpreted numerically because descriptive statistics are not provided. The effect is negative, which implies that more centrality reduces shortages. The effect is significant at 5% level (the t-value is -0.0173/0.0036 = -4.80, or 4.80 in absolute value, and the critical value is 1.96, see statistical table G.2 at the end of the question sheet). The estimate R-squared implies that 12.5% of the variation in shortages is explained by the model.

e) An adviser suggests that you should include the number of students in the model. More teachers do not imply reduced job pressure if also the number of students increases. It is the number of teachers given the number students that should matter. The adviser suggests that you estimate the model

(2) $$\mathbf{Short_i = \beta_0 + \beta_1 \log Teachers_i + \beta_2 Central_i + \beta_3 \log Students_i + u_i}$$

The results are reported in column (2) in the table. Comment on the findings.

The inclusion of the number of students changes the regression results. Thus, it must be concluded that the first model (column (1)) has a problem with omitted variabel. The effect of *logStudents* is negative and significant (t-value = 2.46). The effects of the other variables changes. The effect of *logTeachers* becomes much larger and significant (t-value = 3.76). The estimate implies that increased number of teachers by 10%, given the number of students, increases shortages by 0.779 percentage points. The effect of centrality gets smaller in absolute value and insignificant. R-squared increases by definition, but also R-squared adjusted increases. The latter follows from the fact that the new variable (*logStudents*) have a significant effect.

f) The models in column (4) and (5) in Table 3 have imposed restrictions on equation (2). What are the restrictions? Test separately whether the model in column (4) and the model in column (5) are valid restrictions of the model in column (2).

The model in column (3) is a re-formulation of equation (2). It is

$$\begin{aligned} \text{Short}_i &= \beta_0 + \beta_1 \log \text{Teachers}_i + \beta_2 \text{Central}_i + \beta_3 \log \text{Students}_i + u_i \\ &= \beta_0 - \beta_3 (\log \text{Teachers}_i - \log \text{Students}) + \beta_2 \text{Central}_i + (\beta_1 + \beta_3) \log \text{Teachers}_i + u_i \end{aligned}$$

Thus, the coefficient in front of *(logTeachers – logStudents*) in column (3) ($-\beta_3$) is the same as in front of *logStudents* in column (2), but with opposite sign. Because the models in column (2) and (3) are different formulations of the same model, the parameter for *Central* and R-squared are identical in column (2) og (3). The model in column (4) has imposed the restriction on the model in column (3) than there is no effect of *logTeachers*, i.e., $\beta_1 = -\beta_3$. That is, the effects of *logTeachers* and *logStudents* are equal with opposite signs. The test of the restriction is a test of whether *logTeaches* is significant in column (3). We cannot reject that the restriction is valid because the t-value is below the critical level (t-value is 0.30/0.45 = 0.67). The model in column (5) additionally impose the restriction that $\beta_2 = 0$. We have multiple restrictions ($\beta_1 = -\beta_3$ and $\beta_2 = 0$), and a t-test cannot be used. It is possible to use an F-test that compare the sum of squared residuals between the models in column (3) and (5). However, the sum of squared residuals is not provided. Thus, we have to use a the R-squared form of the F statistic (chapter 4-5c in the textbook). It follows that

$$F = \frac{(R^2_{ur} - R^2_r)}{1 - R^2_{ur}} * \frac{n-k-1}{q}$$

where R^2_{ur} is the R-squared from the unrestricted model (column 3), R^2_r is the R-squared from the restricted model (column 5), q is the number of restrictions (we have 2 restrictions) and (n-k-1) is the degrees of freedom (we have n=412 and k=3). It follows that $F \sim F_{q,n-k-1}$ and

$$F = \frac{0.1533 - 0.1498}{1 - 0.1533} * \frac{408}{2} = 0.84$$

It follows from table G.3b at the end of the question sheet that the critical value at 10% level for $F \sim F_{2,408} = 2.30$. The null hypothesis is not rejected and the restriction is valid.

g) You are informed that you can get data for all years from 2010 to 2019. Can you use this extended data set to improve the credibility of your results? How can you specify the empirical model?

It might be omitted variables in the model above, making the assumption iv) (see question a)) invalid. With more information, it is possible to control for unobserved factors. It is possible to estimate a fixed effect model

$$Short_{it} = \beta_0 + \beta_1 \log Teachers_{it} + \beta_2 Central_i + \beta_3 \log Students_{it} + a_i + u_{it}$$

where subscript t denotes year and a_i is a municipality fixed effect, e.g., a set of dummy variables for each municipality. a_i captures all variations that is constant over time for the municipality. Thus, the model control for more factors than the simple cross-section equation (1) and the results will be more credible. The identification of the effect of the variable Central can be discussed. Inclusion of time specific effects (dummy variables for years) can be discussed. When it comes to empirical specification, it is expected that the specification above is presented, and it is also relevant to discuss the within-specification (textbook ch. 14-1) The first differencing method (ch. 13-5) might also be presented. The framework estimates causal effects if there are no omitted variables.

h) You are also informed that some municipalities increased the number of teachers in 2017 because they received increased grants from the national government. Some municipalities experienced a substantial change from this year. Can you use this information to improve the credibility of your results? How will you specify the empirical model?

It is a change in our empirical period affecting some municipalities and not others. The difference-in-differences method can then be applied to estimate causal effects. It is necessary to include the terms treated municipalities (or treatment group) and control group. The idea is that the difference between the treatment group and the control group changes after the new policy in 2017 compared to the situation before the policy. It is only expected a verbal presentation. It might, however, be useful to present a formal model. This is not straightforward because our model includes the variable *logTeachers* and not a policy variable formulated as a dummy variable. One way to handle the issue is to define a policy variably, say P_{it}, that is equal to one for the control group after the policy and zero otherwise and estimate

$$\Delta Short_{it} = \beta_0 + \beta_1 P_{it} + \beta_3 \Delta \log Students_{it} + \Delta u_{it}$$

where Δ denotes the first difference. It is also relevant to discuss an instrumental variable approach, where grants is an instrument for the number of teachers.

Table 3. Estimation results. Estimated standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)
logTeachers	0.37 (0.46)	7.79 (2.07)	0.30 (0.45)	-	-
logStudents	-	-7.49 (3.04)	-	-	-
Central	-0.0173 (0.0036)	-0.0060 (0.0047)	-0.0060 (0.0047)	-0.0040 (0.0035)	-
logTeachers – logStudents	-	-	7.49 (2.04)	7.55 (2.04)	9.45 (1.11)
Constant	14.94 (1.31)	24.85 (2.99)	24.85 (2.99)	24.85 (2.99)	26.61 (2.54)
R-squared	0.1252	0.1533	0.1533	0.1524	0.1498
R-squared adjusted	0.1210	0.1470	0.1470	0.1482	0.1477
Observations	412	412	412	412	412

Statistical tables

TABLE G.2

Critical Values of the *t* Distribution

		Significance Level				
	1-Tailed:	.10	.05	.025	.01	.005
	2-Tailed:	.20	.10	.05	.02	.01
Degrees of Freedom	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical for a two-tailed test with large (> 120) *df* is 1.96.

Source: This table was generated using the Stata® function invt.

TABLE G.3a

10% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Example: The 10% critical value for numerator $df = 2$ and denominator $df = 40$ is 2.44.
Source: This table was generated using the Stata® function invfprob.

TABLE G.3b

5% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator $df\ (\infty)$ is 2.37.
Source: This table was generated using the Stata® function invfprob.