ggt Examples

Examples

In this section, we present an example to show how the data should be arranged in order to use `ggt`.

Assume we are interested in hospital quality. We have data on 300 patients and 8 hospitals. The individual patient variables include the following: the mortality measure ("mortality"), the hospital choice variable ("hosp_choice"), and an illness severity measure ("severity"). Additionally, we have two variables "dist" and "dist2" representing the distance from each patient to each hospital along with its square (normalized to have similar scales, necessary since the priors are the same). We also have hospital characteristic variables, "hosp_size" and "hosp_ownership". The categories for hosp_size are "small" and "large", and the categories for "hosp_ownership" are "public" and "private".

The individual patient ID variable is called "indnumber" and the hospital ID variable is called "hospnum". In the Stata dataset, there should be an observation for each individual-hospital pair, even if the individual did not choose that hospital. For example, with 300 patients and 8 hospitals, we have 300*8=2400 observations in the data. The table below shows the structure of the data for the first 2 patients. You can see that individual 1 went to hospital 7 and died while patient number 2 went to hospital 3 and did not die. The severity measure is constant within an individual while the distance and distance^2 measures differ for each patient-hospital pair. Additionally, notice the hospital characteristics are constant within hospitals, e.g., hosp_size and hosp_ownership is always "small", "public" for the row in which hospnum==1.

indnumber	hospnum	hosp_choice	mortality	severity	dist	dist2	hosp_size	hosp_ownership
1	1	0	1	1.549	0.015	0.000	small	public
1	2	0	1	1.549	0.250	0.013	large	public
1	3	0	1	1.549	0.259	0.014	large	public
1	4	0	1	1.549	0.080	0.001	large	private
1	5	0	1	1.549	0.097	0.002	large	public
1	6	0	1	1.549	0.160	0.005	large	public
1	7	1	1	1.549	0.459	0.042	small	private
1	8	0	1	1.549	0.491	0.048	small	public
2	1	0	0	0.723	0.052	0.001	small	public
2	2	0	0	0.723	0.162	0.005	large	public
2	3	1	0	0.723	0.067	0.001	large	public
2	4	0	0	0.723	0.097	0.002	large	private
2	5	0	0	0.723	0.187	0.007	large	public
2	6	0	0	0.723	0.019	0.000	large	public
2	7	0	0	0.723	0.110	0.002	small	private
2	8	0	0	0.723	0.058	0.001	small	public
3	1	0	1	2.684	0.142	0.004	small	public
3	2	0	1	2.684	0.070	0.001	large	public

Example 1:

If we want to see the hospital quality measures, β, using all the default settings, we would simply type the command:

```
ggt, outcomevar(mortality) orgchoice(hosp_choice) indID(indnumber) orgID(hospnum)
choicechar(dist dist2)
```

This will apply the selection model with using dist and dist2 as the choice characteristics. Since we did not specify indchar option, the code will assume only a constant and the hospital choice for the individual probit model. Additionally, since we did not specify orgchar, the code will assume no correlation across hospitals via hospital size or ownership. The sampling algorithm will assume the default prior variance options and number of iterations.

The output on the screen will be the summary statistics for the estimated β draws via the MCMC Gibbs sampler. The output for this example is shown below with "q_n" represented the quality for hospital ID, n. Notice that the number of observations is 900- this comes from the default 100,000 iterations, saving only every 100th draw, and deleting the first 10,000 draws as burn-in.

Variable	Obs	Mean	Std. Dev.	Min	Max
q_8	900	-.5500745	.3616526	-2.2093	.817451
q_7	900	.4810402	.2885439	-.361103	1.73498
q_6	900	.0461915	.2821159	-.803522	1.06977
q_5	900	.0610291	.2807342	-1.14918	1.13188
q_4	900	.0410323	.2765834	-.83994	1.01772
q_3	900	-.0477848	.2811626	-1.13624	.952603
q_2	900	.1456833	.3037255	-.930358	1.15258
q_1	900	-.2295427	.2927853	-1.42508	.574466

Note: The code may take several minutes or several hours to complete running depending on the nature of the data. Once the code is complete, the word "complete" will display on the Stata screen. If the code does not complete or Stata simply quits, this is likely due to an error with the prior variance specifications which are not compatible with the data. We suggest trying to call the program again using different prior variance values. If you receive an unspecified error code, please contact the program authors.

Example 2:

Now suppose we want to include the severity measure in the morality equation, and we also want to allow hospital correlation based on size and ownership. Additionally, we want to rescale the prior variances based on the structure of the data. Specifically, we want the prior variance of alpha to be 5, the prior variance of gamma to be 3, selection term for delta to be 0.1, and the parameters for the hyperpriors to be 1 and 5. Finally, we want to save the draws for each of the parameters in a csv file to the directory.

To do this, we would type the command: `ggt, outcomevar(mortality) orgchoice(hosp_choice) indID(indnumber) orgID(hospnum) choicechar(dist dist2) indchar(severity) orgchar(hosp_size hosp_ownership) alphapriorvar(5) gammapriorvar(3) deltapriorvar(.1) priortau(1,5) savedraws`

The output in this case is now:

Variable	Obs	Mean	Std. Dev.	Min	Max
q_8	900	-1.309253	.8931668	-5.08144	1.324964
q_7	900	.4705184	.7550874	-2.787393	2.860252
q_6	900	-.0211944	.7737148	-2.685479	2.600317
q_5	900	-.1552113	.780752	-2.665826	2.67078
q_4	900	.1045821	.7621006	-2.425218	2.938717
q_3	900	-.4967206	.7883454	-3.723473	2.227839
q_2	900	-.0616016	.7767661	-2.638984	2.560112
q_1	900	-.8505865	.789035	-3.939464	1.607428

Additionally, a file called "temp_GGT_output.csv" is saved in the directory. A screenshot of the first 11 rows and 6 columns is shown below.

The official column names are : iter, tau0, tau1, tau2, beta_orgatt1_type1, beta_orgatt1_type2, beta_orgatt2_type1, beta_orgatt2_type2, beta_orgatt3_type1, beta_orgatt3_type2, beta_orgatt3_type3, beta_orgatt3_type4, beta_orgatt3_type5, beta_orgatt3_type6, beta_orgatt3_type7, beta_orgatt3_type8, gamma1, gamma2, alpha1, alpha2, delta1, delta2, delta3, delta4, delta5, delta6, delta7.

-iter: indicates the Gibbs Sampler iteration.
-tau0, tau1, and tau2: the hyperprior draws for variances of hosp_size, hosp_ownership, and hospital organization dummies respectively.
-beta_orgattN_typeM: the β coefficient draws for the dummy variable indicating the Nth specified organization characteristic variable and the Mth category for that variable (where categories are ordered numerically in the case where the variable is string or factor).
Note: In this case, since we specified 2 organization characteristics, beta_orgatt3_typej corresponds to the β coefficient on the dummy variable for hospital j.
-gamma1, gamma2: the γ estimate draws for the coefficient on a constant and the severity measure (respectively) in the outcome probit equation.
-alpha1, alpha2: the α estimate draws for the coefficient on dist and dist2 (respectively) in the organization choice equation.
-delta1-delta7: the δ estimate draws in the selection equation.

iter	tau0	tau1	tau2	beta_orgatt1_type1	beta_orgatt1_type2	beta_orgatt2_type1	beta_orgatt2_type2
100	0.358642	0.220737	0.23951	-0.501334	-1.02871	-0.0908698	-0.477842
200	0.171539	0.329673	0.176983	-0.313839	-0.10947	-0.108144	-0.555235
300	0.293144	0.347339	0.149375	0.217793	0.632413	0.589582	0.209514
400	0.661389	0.318063	0.321596	-0.758161	-1.31175	0.349496	-0.489751
500	0.291911	0.4343	0.51281	0.190255	-0.253484	-0.540287	-0.695708
600	0.26623	0.240774	0.255381	-0.236942	-0.227226	0.673004	-0.508051
700	0.379155	0.314184	0.441279	0.193889	-0.0906312	-0.528254	-0.960122
800	0.216186	0.518378	0.298987	0.414226	0.0352562	1.8174	0.389026
900	1.20827	0.338563	0.126843	1.76718	1.14654	-0.063195	-0.616714
1000	0.692699	0.482613	0.154529	0.592075	-0.144183	0.711893	-0.134483

Example 3:
Finally, suppose we wish to compare the results to the case where we do not apply the selection correction. In this case, the program simply estimates equation (1) in GGT. We can still specify all the options, but the code will only use those that are necessary. e.g., since the nonselection model assumes that δ=0, then specifying deltapriorvar is unnecessary. To run this model, we need to specify the "nonselection" option.

Note: Even though the equation we wish to estimate does not depend on patient-organization choice characteristics, the code will still require choice characteristics in its estimation of α.

```
ggt, outcomevar(mortality) orgchoice(hosp_choice) indID(indnumber) orgID(hospnum)
choicechar(dist dist2) indchar(severity) orgchar(hosp_size hosp_ownership)
alphapriorvar(5) gammapriorvar(3) priortau(1,5) noselection
```

The output for this scenario is as follows:

Variable	Obs	Mean	Std. Dev.	Min	Max
q_8	900	-.7867516	.5794054	-3.822582	1.01066
q_7	900	.2010573	.5644178	-2.832025	2.09774
q_6	900	-.0247391	.5788897	-2.722792	1.938399
q_5	900	-.1067564	.5616703	-3.02365	1.807913
q_4	900	.0623483	.5506349	-2.436699	1.650944
q_3	900	-.3221626	.5825404	-2.83032	1.668997
q_2	900	.0484907	.5699824	-2.7295	2.280766
q_1	900	-.6415367	.5895134	-3.490691	1.766314