An interpretation and implementation of the Theil–Goldberger ‘mixed’ estimator

Christopher F Baum

Boston College and DIW Berlin

Stata Conference 2011, Chicago

Linear regression with non-sample information

In social science research, we often have some *non-sample information* from prior studies regarding plausible parameter values or intervals. We could follow the classical statistical approach, producing point and interval estimates from an estimated regression model and testing whether those estimates are in line with those derived from similar models and/or other data.

As an alternative, if we were of a Bayesian persuasion, we might choose to incorporate this non-sample information explicitly into the estimation problem by means of an informative prior.

What I discuss today is a middle ground between those two approaches, where we use classical statistical techniques but impose constraints—either exact and stochastic—upon the estimation problem.

As an alternative, if we were of a Bayesian persuasion, we might choose to incorporate this non-sample information explicitly into the estimation problem by means of an informative prior.

What I discuss today is a middle ground between those two approaches, where we use classical statistical techniques but impose constraints—either exact and stochastic—upon the estimation problem.

Introducing exact non-sample information

As a starting point, consider the estimation of a linear regression subject to one or more exact linear constraints on the parameter vector. This can be viewed as

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

subject to

$$\mathbf{R}\beta = \mathbf{r}$$

where $\mathbf{R}$ is $J \times K$ and $\mathbf{r}$ is $J \times 1$.

This *constrained least squares* estimator is that implemented by Stata's `cnsreg` command.

Introducing exact non-sample information

As a starting point, consider the estimation of a linear regression subject to one or more exact linear constraints on the parameter vector. This can be viewed as

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

subject to

$$\mathbf{R}\beta = \mathbf{r}$$

where **R** is $J \times K$ and **r** is $J \times 1$.

This *constrained least squares* estimator is that implemented by Stata's `cnsreg` command.

Each row of the constraint system imposes one restriction on the parameter vector, reducing its effective dimensionality from the unconstrained regression. The constraints, which are often adding-up conditions or equality constraints, force the regression model to the suboptimum defined by the constrained system. The constraints must be linearly independent and consistent with each other.

From a textbook treatment of this Lagrangian optimization problem, we can write the estimated parameter vector $\mathbf{b}_{RLS}$ as:

$$\mathbf{b}_{RLS} = \mathbf{b}_{OLS} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}_{OLS} - \mathbf{r})$$

where $\mathbf{b}_{OLS}$ is the vector of unconstrained OLS regression estimates.

Each row of the constraint system imposes one restriction on the parameter vector, reducing its effective dimensionality from the unconstrained regression. The constraints, which are often adding-up conditions or equality constraints, force the regression model to the suboptimum defined by the constrained system. The constraints must be linearly independent and consistent with each other.

From a textbook treatment of this Lagrangian optimization problem, we can write the estimated parameter vector $\mathbf{b}_{RLS}$ as:

$$\mathbf{b}_{RLS} = \mathbf{b}_{OLS} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}_{OLS} - \mathbf{r})$$

where $\mathbf{b}_{OLS}$ is the vector of unconstrained OLS regression estimates.

The restricted least squares (RLS) estimator may be seen as a corrected version of the OLS estimator in which the correction factor for each parameter relates to the magnitude of the $J \times 1$ discrepancy vector $\mathbf{m} = (\mathbf{Rb} - \mathbf{r})$.

To compare the unconstrained (OLS) and constrained (RLS) solutions, we may form a F–statistic from the expression in the difference of sums of squared residuals:

$$e_0' e_0 - e' e = (\mathbf{Rb} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} (\mathbf{Rb} - \mathbf{r})$$

where e_0 and e are, respectively, the RLS and OLS residuals.

The restricted least squares (RLS) estimator may be seen as a corrected version of the OLS estimator in which the correction factor for each parameter relates to the magnitude of the $J \times 1$ discrepancy vector $\mathbf{m} = (\mathbf{Rb} - \mathbf{r})$.

To compare the unconstrained (OLS) and constrained (RLS) solutions, we may form a F–statistic from the expression in the difference of sums of squared residuals:

$$e_0' e_0 - e' e = (\mathbf{Rb} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{Rb} - \mathbf{r})$$

where e_0 and e are, respectively, the RLS and OLS residuals.

This gives rise to the F statistic

$$F[J, n-K] = \frac{(e_0' e_0 - e' e)/J}{e' e/(n-K)}$$

which can be transformed into

$$F[J, n-K] = \frac{(R^2 - R_0^2)/J}{(1-R^2)/(n-K)}$$

In this context, the effect of the J restrictions on the parameter vector may be viewed as either the loss in the least squares criterion or the reduction in R^2 caused by the restrictions. The numerator of either expression is non-negative, as imposition of the restrictions cannot increase R^2 nor can it decrease the sum of squared residuals.

This gives rise to the F statistic

$$F[J, n-K] = \frac{(e_0' e_0 - e' e)/J}{e' e/(n-K)}$$

which can be transformed into

$$F[J, n-K] = \frac{(R^2 - R_0^2)/J}{(1-R^2)/(n-K)}$$

In this context, the effect of the J restrictions on the parameter vector may be viewed as either the loss in the least squares criterion or the reduction in R^2 caused by the restrictions. The numerator of either expression is non-negative, as imposition of the restrictions cannot increase R^2 nor can it decrease the sum of squared residuals.

The limitation of this approach should be evident: constrained least squares allows us to impose non-sample information on the estimation process, but that is an 'all or nothing' choice. The constraints that are imposed are imposed with certainty, as if we are absolutely certain of their validity.

Although we can compare and formally test these constrained estimates to their unconstrained counterparts, we must either utilize the non-sample information or discard it. There is no middle ground.

The limitation of this approach should be evident: constrained least squares allows us to impose non-sample information on the estimation process, but that is an ‘all or nothing’ choice. The constraints that are imposed are imposed with certainty, as if we are absolutely certain of their validity.

Although we can compare and formally test these constrained estimates to their unconstrained counterparts, we must either utilize the non-sample information or discard it. There is no middle ground.

Introducing stochastic non-sample information

As an alternative to the exact non-sample information ($\mathbf{R}\beta = \mathbf{r}$), we might have *stochastic* non-sample information of the form:

$$\mathbf{r} = \mathbf{R}\beta + \upsilon,$$

where $\mathbf{R}, \mathbf{r}$ are defined as before, and υ is a $J \times 1$ unobservable, normally distributed random vector with mean δ and covariance matrix Φ, with Φ known. In this context, δ measures the degree to which the restrictions embodied in $\mathbf{R}, \mathbf{r}$ fail to hold in the population model. If they are thought to hold, $\delta = 0$.

Stochastic non-sample information may take the form of the parameter values obtained from a meta-analysis of our model, and some measure of the degree of precision of those meta-estimates. In this discussion, we will consider a limited form of information: that embodied by parameter values and measures of precision that we are willing to attribute to those values.

For simplicity, the measures of precision may be expressed as standard errors, reflecting the confidence that we are willing to place on each parameter's non-sample value. In this rubric, Φ is taken to be a positive semidefinite diagonal matrix. This allows non-sample information to be present for a subset of the regression parameters.

Stochastic non-sample information may take the form of the parameter values obtained from a meta-analysis of our model, and some measure of the degree of precision of those meta-estimates. In this discussion, we will consider a limited form of information: that embodied by parameter values and measures of precision that we are willing to attribute to those values.

For simplicity, the measures of precision may be expressed as standard errors, reflecting the confidence that we are willing to place on each parameter's non-sample value. In this rubric, Φ is taken to be a positive semidefinite diagonal matrix. This allows non-sample information to be present for a subset of the regression parameters.

Historical context

Before presenting the implementation of this estimator, let us consider its historical context. It is known as the Theil–Goldberger *mixed* (TGM) estimator, as introduced in their 1961 paper ‘On pure and mixed statistical estimation in economics’ (*Int. Ec. Rev.*) and Theil’s 1963 paper ‘On the use of incomplete prior information in regression analysis’ (*JASA*).

The authors’ use of *mixed* in this context is appropriately descriptive, as the estimator they define indeed mixes sample and non-sample information in a generalized least squares sense. Unfortunately, the term nowadays is commonly applied to a quite different set of estimation techniques (e.g., `xtmixed` and `gllamm` in Stata).

Historical context

Before presenting the implementation of this estimator, let us consider its historical context. It is known as the Theil–Goldberger *mixed* (TGM) estimator, as introduced in their 1961 paper ‘On pure and mixed statistical estimation in economics’ (*Int. Ec. Rev.*) and Theil’s 1963 paper ‘On the use of incomplete prior information in regression analysis’ (*JASA*).

The authors’ use of *mixed* in this context is appropriately descriptive, as the estimator they define indeed mixes sample and non-sample information in a generalized least squares sense. Unfortunately, the term nowadays is commonly applied to a quite different set of estimation techniques (e.g., `xtmixed` and `gllamm` in Stata).

Interestingly, Theil and Goldberger point out that a limited form of their suggested strategy of mixing sample and non-sample information was actually proposed by Durbin in a 1953 paper in *JASA*, 'A note on regression when there is extraneous information about one of the coefficients'. They also cite Richard Stone's classic text on consumer expenditure as proposing a maximum-likelihood version of the same routine.

Theil and Goldberger cast the problem as one of generalized least squares, in which the linear statistical model contains both sample and non-sample information:

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \epsilon \\ \upsilon \end{bmatrix}$$

where the vector of errors $(\epsilon', \upsilon')'$ is multivariate Normal with covariance matrix:

$$\begin{bmatrix} \Omega & 0 \\ 0 & \Phi \end{bmatrix}$$

The Theil–Goldberger mixed estimator for this model may be written as:

$$b_{TG} = (\mathbf{X}'\Omega^{-1}\mathbf{X} + \mathbf{R}'\Phi^{-1}\mathbf{R})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{y} + \mathbf{R}'\Phi^{-1}r)$$

with covariance matrix:

$$VCE(b_{TG}) = [\mathbf{X}'\Omega^{-1}\mathbf{X} + \mathbf{R}'\Phi^{-1}\mathbf{R}]^{-1}$$

Under the assumption of *i.i.d.* errors, $\Omega = \sigma^2 I_T$, and σ^2 can be replaced with its consistent OLS estimate s^2 when computing Ω^{-1}.

In Theil's 1963 paper, he develops the mixed estimator in order to incorporate two types of non-sample information: *statistical* information, in which prior research has produced plausible values for coefficients, and *a priori* information, such as that resulting from inequality constraints.

For the latter, he suggests that by placing appropriately chosen measures of precision on coefficients, one can virtually guarantee that the resulting estimate lies in the appropriate range. For instance, if the coefficient β_1 is thought to almost surely lie between 0 and 1, and probably between 0.25 and 0.75, with an implied standard error of $\frac{1}{4}$, we could specify that

$$0.5 = \beta_1 + \upsilon; \quad E\upsilon = 0; \quad E\upsilon^2 = \frac{1}{16}.$$

In Theil's 1963 paper, he develops the mixed estimator in order to incorporate two types of non-sample information: *statistical* information, in which prior research has produced plausible values for coefficients, and *a priori* information, such as that resulting from inequality constraints.

For the latter, he suggests that by placing appropriately chosen measures of precision on coefficients, one can virtually guarantee that the resulting estimate lies in the appropriate range. For instance, if the coefficient β_1 is thought to almost surely lie between 0 and 1, and probably between 0.25 and 0.75, with an implied standard error of $\frac{1}{4}$, we could specify that

$$0.5 = \beta_1 + \upsilon; \quad E\upsilon = 0; \quad E\upsilon^2 = \frac{1}{16}.$$

In their 1961 paper, Theil and Goldberger suggest that the estimator may be applied to linear combinations of coefficients about which there is some *a priori* knowledge; for instance, in economics, constant returns to scale (CRTS) in production requires that elasticities of a Cobb–Douglas function sum to unity.

They also illustrate that this technique may also be applied to two-stage least squares (2SLS) estimates, and outline a procedure by which the σ^2 estimate used to produce the covariance matrix of the estimated parameters may be refined by iteration to convergence.

The implementation I present below has not yet been extended for these three enhancements.

In their 1961 paper, Theil and Goldberger suggest that the estimator may be applied to linear combinations of coefficients about which there is some *a priori* knowledge; for instance, in economics, constant returns to scale (CRTS) in production requires that elasticities of a Cobb–Douglas function sum to unity.

They also illustrate that this technique may also be applied to two-stage least squares (2SLS) estimates, and outline a procedure by which the σ^2 estimate used to produce the covariance matrix of the estimated parameters may be refined by iteration to convergence.

The implementation I present below has not yet been extended for these three enhancements.

In their 1961 paper, Theil and Goldberger suggest that the estimator may be applied to linear combinations of coefficients about which there is some *a priori* knowledge; for instance, in economics, constant returns to scale (CRTS) in production requires that elasticities of a Cobb–Douglas function sum to unity.

They also illustrate that this technique may also be applied to two-stage least squares (2SLS) estimates, and outline a procedure by which the σ^2 estimate used to produce the covariance matrix of the estimated parameters may be refined by iteration to convergence.

The implementation I present below has not yet been extended for these three enhancements.

In his 1963 paper, Theil proposes a formal test of the *compatibility* of prior and sample information. Under the null hypothesis that the two sets of information are in agreement, we have two estimators of the vector $\mathbf{R}\beta$: the prior estimator **R** and the OLS estimator b_{OLS}. Under the assumption that υ has zero mean and is normally distributed, he derives the test statistic:

$$\hat{\gamma} = (\mathbf{r} - \mathbf{R}\mathbf{b}_{OLS})' \left[s^2 [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' + \Phi \right]^{-1} (\mathbf{r} - \mathbf{R}\mathbf{b}_{OLS})$$

which he shows is distributed as χ^2 under the null hypothesis, with degrees of freedom equal to the rank of Φ.

Conway and Mittelhammer (*Stud. Ec. Analysis*, 1986, p. 89) point out that a rejection of the null is a rejection of the unbiasedness of the prior information: that is, that $E\upsilon = 0$.

In his 1963 paper, Theil proposes a formal test of the *compatibility* of prior and sample information. Under the null hypothesis that the two sets of information are in agreement, we have two estimators of the vector $\mathbf{R}\beta$: the prior estimator $\mathbf{R}$ and the OLS estimator b_{OLS}. Under the assumption that v has zero mean and is normally distributed, he derives the test statistic:

$$\hat{\gamma} = (\mathbf{r} - \mathbf{R}\mathbf{b}_{OLS})' \left[s^2 [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' + \Phi \right]^{-1} (\mathbf{r} - \mathbf{R}\mathbf{b}_{OLS})$$

which he shows is distributed as χ^2 under the null hypothesis, with degrees of freedom equal to the rank of Φ.

Conway and Mittelhammer (*Stud. Ec. Analysis*, 1986, p. 89) point out that a rejection of the null is a rejection of the unbiasedness of the prior information: that is, that $Ev = 0$.

In the same paper, Theil proposes scalar measures of the shares of prior sample information in the posterior precision of the TGM estimates. To what degree are the mixed estimates merely reflecting our subjective beliefs, as expressed by the priors? He shows that

$$\theta_S = \frac{1}{K} \operatorname{tr} s^{-2}\mathbf{X}'\mathbf{X} \left(s^{-2}\mathbf{X}'\mathbf{X} + \mathbf{R}'\Psi^{-1}\mathbf{R} \right)^{-1}$$

expresses the share due to sample information, while $\theta_P = 1 - \theta_S$ expresses the share due to prior information.

In 1980, V. K. Srivastava published ‘Estimation of linear single-equation and simultaneous-equation models under stochastic linear constraints: An annotated bibliography’ (*Intl. Stat. Rev.*). After more than two decades of research in this area, a modest number of papers are listed, most of them focusing on the econometric theory of the mixed estimator rather than its practical application.

One notable annotation: that of Swamy and Mehta (*JASA*, 1969), summarized as ‘Assuming the disturbances to follow a normal probability law, it is shown that the mixed estimator is unbiased with a finite variance-covariance matrix and the gain in efficiency over the least squares estimator ignoring the restrictions may be substantial in small samples.’ (p. 81)

In 1980, V. K. Srivastava published ‘Estimation of linear single-equation and simultaneous-equation models under stochastic linear constraints: An annotated bibliography’ (*Intl. Stat. Rev.*). After more than two decades of research in this area, a modest number of papers are listed, most of them focusing on the econometric theory of the mixed estimator rather than its practical application.

One notable annotation: that of Swamy and Mehta (*JASA*, 1969), summarized as ‘Assuming the disturbances to follow a normal probability law, it is shown that the mixed estimator is unbiased with a finite variance-covariance matrix and the gain in efficiency over the least squares estimator ignoring the restrictions may be substantial in small samples.’ (p. 81)

However, in a 1983 *J. Econometrics* article, Swamy and Mehta criticize the Theil–Goldberger approach, arguing that the ‘subjective predictions which Theil employs in his mixed estimation procedure are incorrectly equated to random variables.’ They also claim that the standard errors of this procedure give ‘a spurious sense of precision to the results.’ (pp. 388–389).

The mixed estimation technique has been employed by Mittelhammer and coauthors (*Am. J. Agric. Econ*, 1980, 1988) and, more recently, in a macroeconomic context, by Amato and Gerlach, ‘Modeling the transmission mechanism of monetary policy in emerging market countries using prior information’, *BIS Papers No. 8*, and by Gavin and Kemme, ‘Using extraneous information to analyze monetary policy in transition economics’, *J. Int. Money Fin.*, 2009.

However, in a 1983 *J. Econometrics* article, Swamy and Mehta criticize the Theil–Goldberger approach, arguing that the 'subjective predictions which Theil employs in his mixed estimation procedure are incorrectly equated to random variables.' They also claim that the standard errors of this procedure give 'a spurious sense of precision to the results.' (pp. 388–389).

The mixed estimation technique has been employed by Mittelhammer and coauthors (*Am. J. Agric. Econ*, 1980, 1988) and, more recently, in a macroeconomic context, by Amato and Gerlach, 'Modeling the transmission mechanism of monetary policy in emerging market countries using prior information', *BIS Papers No. 8*, and by Gavin and Kemme, 'Using extraneous information to analyze monetary policy in transition economics', *J. Int. Money Fin.*, 2009.

Amato and Gerlach provide useful intuition for the workings of the TGM, showing that the mixed estimate of the parameter vector β can be written as a matrix weighted average of the prior vector and the OLS estimates:

$$b_{mix} = \mathbf{F}\, b_{prior} + (\mathbf{I} - \mathbf{F})\, b_{OLS}$$

where

$$\mathbf{F} = \left[\Sigma^{-1}_{prior} + \Sigma^{-1}_{OLS}\right]^{-1} \Sigma^{-1}_{prior}$$

so that 'the weight placed on the prior information depends on the confidence the modeler attaches to it.' (p. 266) The constrained least squares estimator sets some of the diagonal elements of **F** to unity, while if we have no prior information about certain coefficients, their respective diagonal elements in **F** will be zero.

The strength of the non-sample information also influences the degree of precision of the mixed estimator. If we estimate a single parameter,

$$s^2_{mix} = \frac{1}{\left[\frac{1}{\sigma^2_{prior}} + \frac{1}{s^2_{OLS}} \right]}$$

From this expression, we can see that (i) as $\sigma^2_{prior} \rightarrow \infty$, $s^2_{mix} \rightarrow s^2_{OLS}$. We may also note that $0 \leq s^2_{mix} \leq s^2_{OLS}$, so that 'the precision of the mixed estimate is at least as high as the precision of the estimate based solely on the data, with the former converging to the latter as the degree of prior uncertainty increases.' (Amato & Gerlach, p. 267)

This also implies that forecasts from the TGM model may be more precise than those from OLS, taking the non-sample information into account.

The strength of the non-sample information also influences the degree of precision of the mixed estimator. If we estimate a single parameter,

$$s^2_{mix} = \frac{1}{\left[\frac{1}{\sigma^2_{prior}} + \frac{1}{s^2_{OLS}}\right]}$$

From this expression, we can see that (i) as $\sigma^2_{prior} \to \infty$, $s^2_{mix} \to s^2_{OLS}$. We may also note that $0 \leq s^2_{mix} \leq s^2_{OLS}$, so that 'the precision of the mixed estimate is at least as high as the precision of the estimate based solely on the data, with the former converging to the latter as the degree of prior uncertainty increases.' (Amato & Gerlach, p. 267)

This also implies that forecasts from the TGM model may be more precise than those from OLS, taking the non-sample information into account.

The strength of the non-sample information also influences the degree of precision of the mixed estimator. If we estimate a single parameter,

$$s^2_{mix} = \frac{1}{\left[\frac{1}{\sigma^2_{prior}} + \frac{1}{s^2_{OLS}}\right]}$$

From this expression, we can see that (i) as $\sigma^2_{prior} \to \infty$, $s^2_{mix} \to s^2_{OLS}$. We may also note that $0 \leq s^2_{mix} \leq s^2_{OLS}$, so that 'the precision of the mixed estimate is at least as high as the precision of the estimate based solely on the data, with the former converging to the latter as the degree of prior uncertainty increases.' (Amato & Gerlach, p. 267)

This also implies that forecasts from the TGM model may be more precise than those from OLS, taking the non-sample information into account.

Implementation

I have developed a first version of the TGM estimator for Stata 11.2+ based upon the analytics given above. The command, `tgmixed`, is an e-class (estimation) command, so that it leaves behind the information needed for common post-estimation commands such as `test`, `lincom`, `predict` and `margins`.

The command syntax:

`tgmixed` *depvar indepvars* `[if` *exp*`] [in` *range*`], prior(`*varname value se...*`)`
`[cov(`*var1 var2 value...*`)] [qui]`

This specifies an OLS regression, with *i.i.d.* errors assumed at present, where you have non-sample information on one or more of the *indepvars* coefficients, given in `prior()`. For each of these coefficients, you specify its variable name, its prior value, and its standard error (se). The optional `cov()` option may be used to specify prior covariances among pairs of coefficients from the *indepvars* list. The `qui` option suppresses the unconstrained OLS estimates.

The command syntax:

`tgmixed` *depvar indepvars* `[if` *exp*`] [in` *range*`], prior(`*varname value se...*`)`
`[cov(`*var1 var2 value...*`)] [qui]`

This specifies an OLS regression, with *i.i.d.* errors assumed at present, where you have non-sample information on one or more of the *indepvars* coefficients, given in `prior()`. For each of these coefficients, you specify its variable name, its prior value, and its standard error (se). The optional `cov()` option may be used to specify prior covariances among pairs of coefficients from the *indepvars* list. The `qui` option suppresses the unconstrained OLS estimates.

By default, for purposes of comparison, the command displays the unconstrained OLS estimates, followed by the parsed values of the `prior()` and, if present, `cov()` options. The TGM estimates are then displayed, along with the Theil compatibility statistic, which gauges the degree of compatibility of sample and non-sample information.

Following estimation, the `predict` command may be used in- or out-of-sample, with available options `xb` (the predicted values of the dependent variable) and `stdp` (the standard errors of prediction.

By default, for purposes of comparison, the command displays the unconstrained OLS estimates, followed by the parsed values of the `prior()` and, if present, `cov()` options. The TGM estimates are then displayed, along with the Theil compatibility statistic, which gauges the degree of compatibility of sample and non-sample information.

Following estimation, the `predict` command may be used in- or out-of-sample, with available options `xb` (the predicted values of the dependent variable) and `stdp` (the standard errors of prediction.

An empirical example

For an illustration of `tgmixed`, I reproduce the computations of Theil (*JASA*, 1963) in which he makes use of 17 annual observations on textile consumption in the Netherlands, 1923–1939. The raw data are provided in the appendix to Theil and Nagar (*JASA*, 1961). The model is

$$\log c_t = \alpha + \beta_1 \log p_t + \beta_2 \log M_t + u_t$$

where c_t is per capita textile consumption, p_t is the deflated price index for textiles, and M_t is real per capita income.

Theil expresses prior beliefs about β_1 and β_2: that they should equal -0.7 and 1.0, respectively, each with a standard error of 0.15. He also specifies that the covariance between the estimated coefficients should be set to -0.01.

These prior values may then be given to `tgmixed` using the `prior()` and optional `cov()` options:

```
tgmixed lconsump lincome lprice, ///
prior(lprice -0.7 0.15 lincome 1 0.15) cov(lprice lincome -0.01)
```

Theil expresses prior beliefs about β_1 and β_2: that they should equal -0.7 and 1.0, respectively, each with a standard error of 0.15. He also specifies that the covariance between the estimated coefficients should be set to -0.01.

These prior values may then be given to `tgmixed` using the `prior()` and optional `cov()` options:

```
tgmixed lconsump lincome lprice, ///
prior(lprice -0.7 0.15 lincome 1 0.15) cov(lprice lincome -0.01)
```

```
. tgmixed lconsump lincome lprice, prior(lprice -0.7 0.15 lincome 1 0.15) cov(l
> price lincome -0.01)
Unconstrained OLS estimates
      Source |       SS       df       MS              Number of obs =      17
-------------+------------------------------           F(  2,    14) =  266.11
       Model |  .097576609     2  .048788305           Prob > F      =  0.0000
    Residual |  .002566775    14  .000183341           R-squared     =  0.9744
-------------+------------------------------           Adj R-squared =  0.9707
       Total |  .100143384    16  .006258962           Root MSE      =  .01354

------------------------------------------------------------------------------
    lconsump |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lincome |   1.143174   .1559813     7.33   0.000     .8086273     1.47772
      lprice |   -.828862   .0361062   -22.96   0.000    -.9063022   -.7514218
       _cons |   1.373925   .3060511     4.49   0.001     .7175102    2.030339
------------------------------------------------------------------------------
```

Note that the data produce very precise estimates of both of the elasticity values, with a point estimate for the income elasticity considerably above unity.

```
Prior coefficient values and standard errors
              1            2

1         lprice     lincome
2           -0.7           1
3           0.15        0.15

Prior covariances
              1

1         lprice
2        lincome
3          -0.01
```

tgmixed reports the prior values placed on the coefficients and on their covariance. Note that you need not express prior beliefs about all of the coefficients.

```
Theil-Goldberger mixed estimates
                                                       Number of obs =         17
                                                       R-squared     =     0.9741
                                                       Root MSE      =     .01361

------------------------------------------------------------------------------
    lconsump |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lincome |   1.089357   .1033892    10.54   0.000     .8676092    1.311105
      lprice |  -.8205463    .034965   -23.47   0.000    -.8955387   -.7455538
       _cons |   1.466644    .203478     7.21   0.000     1.030227    1.903061
------------------------------------------------------------------------------
Theil compatibility statistic =  0.8606            Pr > Chi2(  2) = 0.6503
Shares of posterior precision:   sample info = 0.794   prior info = 0.206
```

The mixed estimates illustrate that the coefficients have been drawn toward the non-sample values: more so for the income coefficient, which had weaker sample information. The R^2 has decreased marginally, while the *RMSE* has increased by less than one per cent.

The mixed point and interval estimates closely match those given in Theil (1963). Likewise, the compatibility statistic of 0.86, with a large p-value, matches his value and indicates that the sample and non-sample information are compatible. Given the precise unconstrained estimates, it is not surprising that the share of sample information in the precision of the mixed estimates is almost 80 percent.

To illustrate the usefulness of the TGM estimator, I conduct a forecast exercise for a model of the change in the log of US real investment spending, 1959Q1–2007Q2, as a function of the change in log US real GDP, the change in the log real wage and the change in the S&P500 stock market index. A constant is included to capture a trend in the level series.

The mixed point and interval estimates closely match those given in Theil (1963). Likewise, the compatibility statistic of 0.86, with a large p-value, matches his value and indicates that the sample and non-sample information are compatible. Given the precise unconstrained estimates, it is not surprising that the share of sample information in the precision of the mixed estimates is almost 80 percent.

To illustrate the usefulness of the TGM estimator, I conduct a forecast exercise for a model of the change in the log of US real investment spending, 1959Q1–2007Q2, as a function of the change in log US real GDP, the change in the log real wage and the change in the S&P500 stock market index. A constant is included to capture a trend in the level series.

```
. reg dlrinv dlrgdp dlrwage dspindex if tin(,2007q2)
      Source |       SS       df       MS              Number of obs =     193
-------------+------------------------------           F(  3,   189) =   62.58
       Model |  .034651338     3  .011550446           Prob > F      =  0.0000
    Residual |   .0348847    189  .000184575           R-squared     =  0.4983
-------------+------------------------------           Adj R-squared =  0.4904
       Total |  .069536038   192  .000362167           Root MSE      =  .01359

------------------------------------------------------------------------------
      dlrinv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dlrgdp |   1.566386   .1234444    12.69   0.000      1.32288    1.809892
     dlrwage |  -.1399312   .1950992    -0.72   0.474    -.5247829    .2449204
    dspindex |   .0541718   .0348277     1.56   0.122    -.0145292    .1228728
       _cons |   -.005029   .0013609    -3.70   0.000    -.0077136   -.0023445
------------------------------------------------------------------------------
```

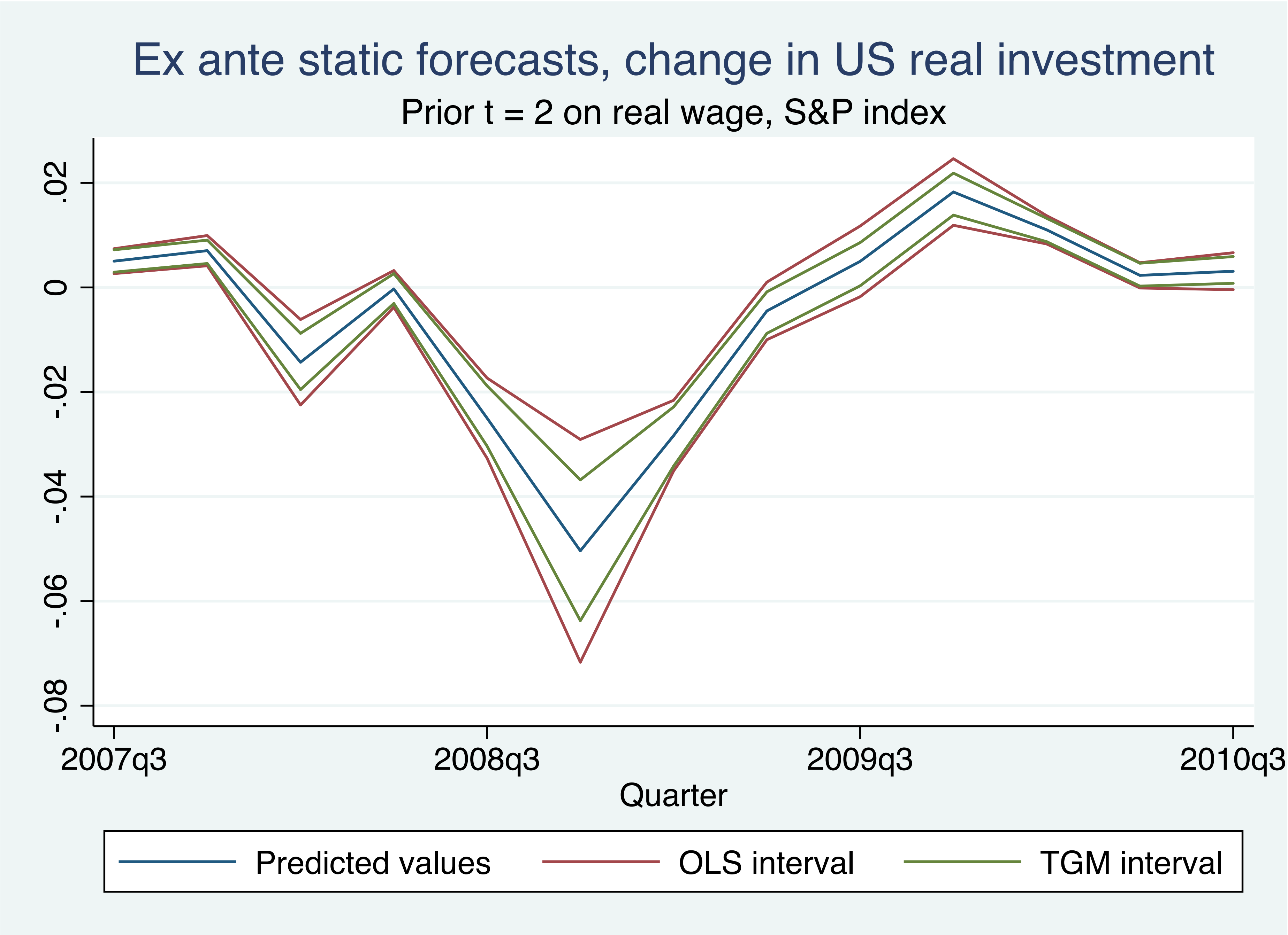
Although the model fits quite well for a first difference specification, the coefficients for the log real wage and stock market index are estimated quite imprecisely. This will weaken the forecast performance of the model.

I apply the TGM estimator, with non-sample point estimates of −0.2 and 0.05 for those variables' coefficients. No prior is specified for the real GDP coefficient. In a first specification, I provide standard errors representing *t*-statistics of 2.0 for each coefficient. The TGM estimates yield:

```
. tgmixed dlrinv dlrgdp dlrwage dspindex if tin(,2007q2), ///
>         prior(dspindex 0.05 0.025 dlrwage -0.2 0.1)
...
Theil-Goldberger mixed estimates
                                                    Number of obs =       193
                                                    R-squared     =    0.4982
                                                    Root MSE      =   .013588

------------------------------------------------------------------------------
      dlrinv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dlrgdp |   1.577946     .116535    13.54   0.000     1.348069    1.807822
     dlrwage |  -.1877885    .0889084    -2.11   0.036    -.3631688   -.0124082
    dspindex |   .0510915    .0202723     2.52   0.013     .0111025    .0910805
       _cons |  -.0050139    .0013579    -3.69   0.000    -.0076924   -.0023354
------------------------------------------------------------------------------
Theil compatibility statistic =   0.0806                 Pr > Chi2(  2) = 0.9605
Shares of posterior precision:   sample info = 0.638   prior info = 0.362
```

Notice that the sample information is responsible for 64 percent of the precision of the estimates, and the compatibility statistic indicates that the non-sample information is reasonably similar to the sample information. We may then produce point and interval static forecasts for the out-of-sample period 2007Q3–2010Q3, and juxtapose them with those from the unconstrained OLS estimates.

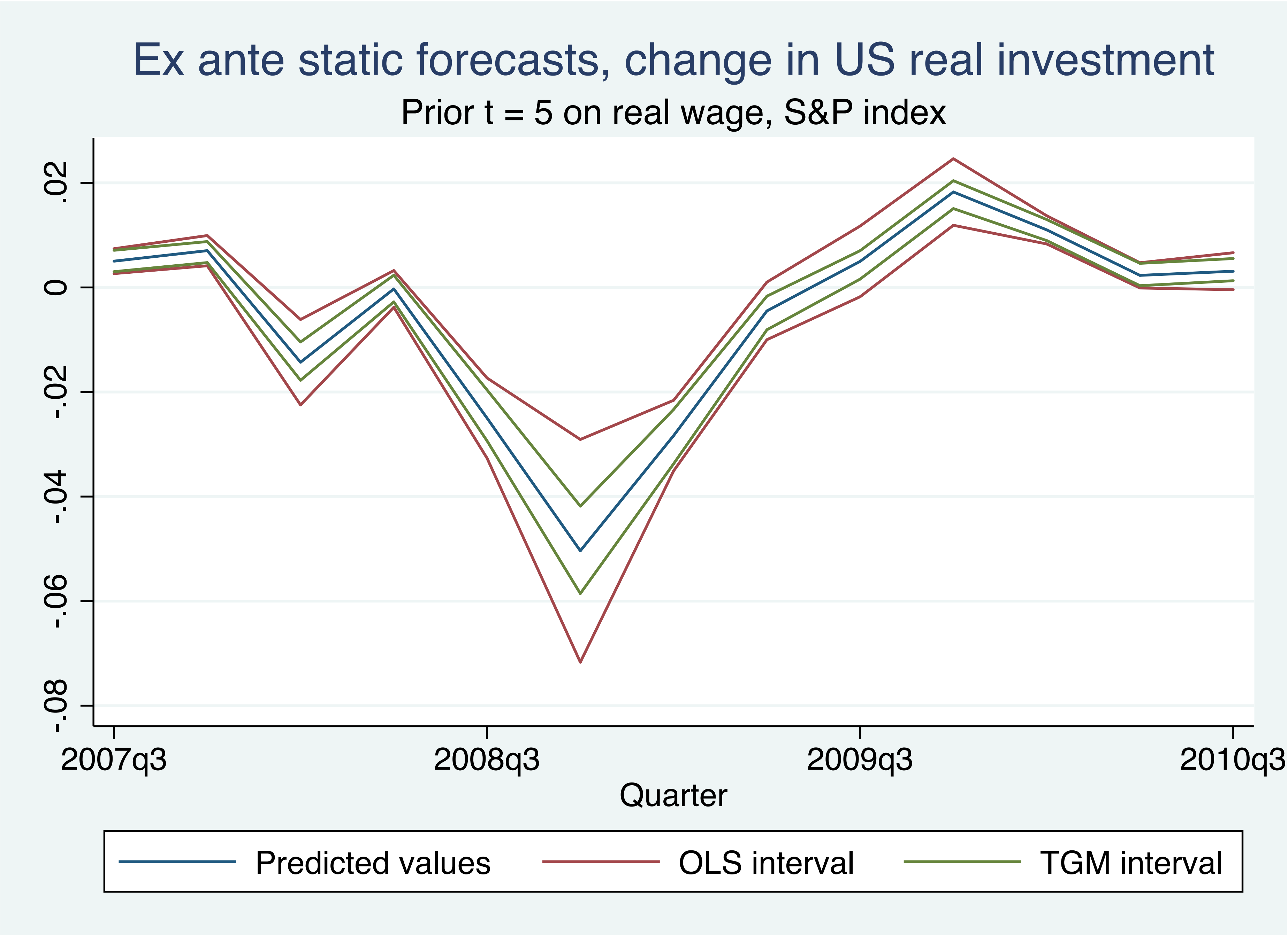

Ex ante static forecasts, change in US real investment
Prior t = 2 on real wage, S&P index
.02
0
-.02
-.04
-.06
-.08
2007q3
2008q3
2009q3
2010q3
Quarter
Predicted values
OLS interval
TGM interval

We may note that the TGM interval estimates are considerably narrower for the downturn at the end of 2008, despite the relatively weak prior. We respecify the prior to reflect *t*-statistics of 5.0 for the two coefficients:

```
. tgmixed dlrinv dlrgdp dlrwage dspindex if tin(,2007q2), ///
>         prior(dspindex 0.05 0.01 dlrwage -0.2 0.04)
...
Theil-Goldberger mixed estimates
                                                       Number of obs =       193
                                                       R-squared     =    0.4981
                                                       Root MSE      =   .013589

------------------------------------------------------------------------------
      dlrinv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dlrgdp |   1.580415   .1149993    13.74   0.000     1.353568    1.807262
     dlrwage |  -.1976498    .039175    -5.05   0.000    -.2749262   -.1203735
    dspindex |   .0502296   .0096069     5.23   0.000     .0312792    .0691801
       _cons |  -.0050101   .0013568    -3.69   0.000    -.0076864   -.0023338
------------------------------------------------------------------------------
Theil compatibility statistic =  0.0978                 Pr > Chi2(  2) = 0.9523
Shares of posterior precision:   sample info = 0.529    prior info = 0.471
```

The sample information is still responsible for 53 percent of the precision of the TGM estimates. The two imprecisely estimated OLS coefficients are now quite close to their prior values. A comparison of the static *ex ante* forecasts versus their OLS counterparts yields:


Ex ante static forecasts, change in US real investment
Prior t = 5 on real wage, S&P index
.02
0
-.02
-.04
-.06
-.08
2007q3
2008q3
2009q3
2010q3
Quarter
Predicted values
OLS interval
TGM interval

Concluding remarks

There are many enhancements that may be added to the `tgmixed` routine, including the ability to handle non-*i.i.d.* errors, time series and factor variables' varlists, general linear constraints on the parameters and the ability to support least squares estimators beyond OLS. Nevertheless, the preliminary routine, just over 200 lines of code—most of it Mata—illustrates the ease of creating a new estimator in Stata.

An important acknowledgement: the syntax parsing code makes use of Ben Jann's Mata routine `mm_posof()`, included in his incredibly useful `moremata` package on SSC. If you use Mata, you should have a copy of `moremata`.