The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*, *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The Stata Journal (2013)
13, Number 4, pp. 836–861

Parametric inference using structural break tests

Zachary L. Flynn
University of Wisconsin–Madison
Madison, WI
zflynn@wisc.edu

Leandro M. Magnusson
University of Western Australia
Crawley, Australia
leandro.magnusson@uwa.edu.au

Abstract. We present methods for testing hypotheses and estimating confidence sets for structural parameters of economic models in the presence of instabilities and breaks of unknown form. These methods constructively explore information generated by changes in the data-generating process to improve the inference of parameters that remain stable over time. The proposed methods are suitable for models cast in the generalized method of moments framework, which makes their application wide. Moreover, they are robust to the presence of weak instruments. The `genstest` command in Stata implements these methods to conduct hypothesis tests and to estimate confidence sets.

Keywords: st0320, genstest, condivreg, ivregress, ivreg2, gmm, qll, generalized method of moments, structural change, weak instruments, hypothesis testing, confidence sets

1 Introduction

We present methods for the inference of parameters in economic models in the presence of instabilities and breaks of unknown form. The main idea behind these methods is to constructively explore exogenous changes in the data-generating process to improve the inference about parameters that are assumed to be stable over time. For example, exogenous changes in the monetary policy induced by the Central Bank affect interest rates. However, if we are interested in parameters that characterize a production function technology, these parameters are not affected by monetary policy. These exogenous variations in the interest rate can be used to improve the inference of such technological parameters. The proposed methods are suitable for models cast in the generalized method of moments (GMM) framework, which makes their application wide. Moreover, they are robust to the presence of weak instruments; that is, we do not assume that the structural parameters are consistently estimated.

Estimation of economic models using GMM departs from a set of moment restrictions, usually derived from the economic theory. Two underlying assumptions of a GMM estimator are that the parameters are stable over time and that they can be consistently estimated using the empirical moment restrictions. These assumptions can be very strong in some economic models. Stock and Watson (1996) and Piehl et al. (2003) report evidence of parameter instability in, respectively, macroeconomic and microeconomic models (for models with identification failure, see Stock, Wright, and Yogo

 st0320

[2002], Kleibergen and Mavroeidis [2009], and references therein). Therefore, we cannot rule out a priori the presence of instabilities and weak instruments.

Structural break tests are considered as a diagnostic test, conducted after estimation. These tests compare the null hypothesis of a stable model (constant parameter) against the alternative of an unstable model (time-varying parameter). Examples include the ones proposed by Andrews (1993) and Elliott and Müller (2006). Moreover, they assume that the parameters are consistently estimated under the null hypothesis. Instead, we perform hypothesis testing of the parameter of interest and later estimate confidence intervals for this parameter by inverting these tests without estimating the parameter in the first place.

The proposed tests are a combination of two (asymptotically) independent statistics. One of them, the S test (see Stock and Wright [2000]), tests the validity of the moment condition. The second statistic tests the stability of such moments. We call these tests the "generalized S" (gen-S) tests because they are an extension of the S test for stable models. An important feature of the tests is that they are identification robust tests; that is, they have the correct size even in the presence of weak instruments.

The `genstest` command performs these new methods. For a given structural parameter of interest θ, the tests test the simple hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta \neq \theta_0$, where θ_0 is a hypothesized value of θ. The `genstest` command also generates $1 - \alpha$ confidence intervals and sets by inverting the new statistical tests. There are no restrictions on how many parameters may be tested by `genstest`. However, `genstest` calculates confidence sets only up to two parameters because these are the most straightforward to graph (a simple `while` loop can generate a confidence set for any number of parameters).

In section 2, we show how changes in the first stage improve the inference of structural parameters in a linear instrumental variable (IV) model and describe the proposed methods. In section 3, we present the general algorithm for implementing the tests. In section 4, we discuss the syntax and options of the postestimation command `genstest`. Finally, in section 5, we provide examples of its use for performing hypothesis tests and constructing confidence intervals and sets.

2 Structural inference under instability

2.1 A stylized IV model

Consider the following simple limited-information IV model

$$\begin{cases} \mathbf{y} = & \mathbf{Y}\theta + \mathbf{u} \\ \mathbf{Y} = & \mathbf{Z}\Pi + \mathbf{v} \end{cases} \tag{1}$$

where $\mathbf{y}$ is a $T \times 1$ vector, $\mathbf{Y}$ is a $T \times k_Y$ matrix of explanatory variables, $\mathbf{Z}$ is a $T \times k_Z$ matrix of instruments, and $\mathbf{u}$ and $\mathbf{v}$ are residuals. We are interested in testing the following assumption about the value of the structural parameter θ:

$$H_0^\theta : \theta = \theta_0 \quad \text{against} \quad H_1^\theta : \theta \neq \theta_0$$

Postmultiplying the second equation in (1) by θ_0 and subtracting from the first equation, we derive

$$\mathbf{y} - \mathbf{Y}\theta_0 = \mathbf{Z}\delta + \mathbf{e} \tag{2}$$

where $\delta = \Pi(\theta - \theta_0)$ and $\mathbf{e} = \mathbf{u} + \mathbf{v}(\theta - \theta_0)$. Therefore, we can test the null hypothesis $H_0^\theta : \theta = \theta_0$ indirectly by testing the assumption

$$H_0^\delta : \delta = 0 \quad \text{against} \quad H_1^\delta : \delta \neq 0$$

on the auxiliary (2). The principle of testing this null hypothesis by testing violations of the moment restrictions $E\{(1/T)\sum_{t=1}^{T} Z_t(y_t - Y_t\theta)\} = 0$ is from Anderson and Rubin (1949); the auxiliary regression representation in (2) is attributed to Chernozhukov and Hansen (2008). From now on, we will refer to this test simply as the S test, the extended version of the Anderson and Rubin (1949) test for GMM models proposed by Stock and Wright (2000).

The S test has the correct size even when the structural parameter θ is not identified; see Stock and Wright (2000). However, when $\Pi \approx 0$, the S test will not reject H_0^δ when H_1^δ is true.[1] If this is the case, confidence sets derived by inverting the S test are unbounded, giving no information about the location of θ.

In the representation (1), Π captures the strength of the instruments Z_t and is assumed to be the unique solution of

$$E\{Z_t'(Y_t - Z_t\Pi)\} = 0 \quad \text{for all } 0 < t \leq T$$

Now, similar to Angrist and Krueger (1995), we will assume that the strength of the instruments might differ in two subsamples. For simplicity, order the observations such that

$$\begin{aligned} E\{Z_t'(Y_t - Z_t\Pi_1)\} = 0 &\quad \text{for all } 0 < t \leq t_b \\ E\{Z_t'(Y_t - Z_t\Pi_2)\} = 0 &\quad \text{for all } t_b < t \leq T \end{aligned}$$

Partition $\mathbf{Z} = (\mathbf{Z}_1' : \mathbf{Z}_2')'$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are $t_b \times k_Z$ and $(T - t_b) \times k_Z$ submatrices of $\mathbf{Z}$ containing observations of the first and second subsamples, respectively. Define $\overline{\mathbf{Z}}_1 = (\mathbf{Z}_1' : 0')'$ and $\overline{\mathbf{Z}}_2 = (0' : \mathbf{Z}_2')'$ so that $\mathbf{Z} = \overline{\mathbf{Z}}_1 + \overline{\mathbf{Z}}_2$. The first-stage equation is rewritten as

$$\mathbf{Y} = \overline{\mathbf{Z}}_1\Pi_1 + \overline{\mathbf{Z}}_2\Pi_2 + \mathbf{v}$$

and the auxiliary regression becomes

$$\mathbf{y} - \mathbf{Y}\theta_0 = \overline{\mathbf{Z}}_1\delta_1 + \overline{\mathbf{Z}}_2\delta_2 + \mathbf{e} \tag{3}$$

1. When $\Pi = 0$, the estimated value of δ will be close to 0, independent of whether $\|\theta - \theta_0\| > 0$.

We restate the null hypothesis as

$$H_0^{\delta_b}: \delta_1 = \delta_2 = 0 \quad \text{against} \quad H_1^{\delta_b}: \delta_1 \neq 0 \text{ or } \delta_2 \neq 0$$

Therefore, a single change in Π doubles the number of instruments for testing θ. Moreover, the S test applied to the auxiliary regression (3) would have more power to reject $H_0^{\delta_b}$ than the S test applied to (2) to reject H_0^{δ}.

The above linear IV model shows that changes in the first-stage reduced-form parameter can improve the inference about the second-stage structural parameter θ, which remains constant over time or cross-section units. However, in practice, we may not know when the change occurs, the magnitude of the change, or the nature of the instability. In the following subsection, we present tests that do not require such knowledge.

2.2 The generalized S tests

Assume that from an economic model, we derived a moment condition of the form

$$E\left\{Z_t' u\left(\mathcal{Y}_t; \theta, \gamma\right)\right\} = 0 \quad \text{for all } 0 < t \leq T \tag{4}$$

where $u(\cdot;\cdot)$ is a one-dimensional real function indexed by the p-dimensional structural parameter vector θ and by the q-dimensional nuisance parameter vector γ, which is always treated as stable under the null hypothesis. $\mathcal{Y}_t$ is a vector of random variables, and Z_t is the $1 \times k_Z$ dimensional row vector of instruments. For simplicity, we denote $u(\mathcal{Y}_t; \theta, \gamma)$ as $u_t(\theta, \gamma)$. We can consider $u_t(\theta, \gamma)$ as the unobserved error term of a regression such that $E\{u(\mathcal{Y}_t; \theta, \gamma) | Z_t\} = 0$. For instance, in the previous section, $u_t(\theta, \gamma) = y_t - Y_t\theta$. Further examples for cross-section models and time-series models are shown in section 5.

The moment restriction in (4) can be restated in terms of full-sample and stability restrictions as, respectively,

$$E\left\{\frac{1}{T}\sum_{t=1}^{T} Z_t' u_t(\theta, \gamma)\right\} = 0 \quad \text{and} \quad E\left\{Z_t' u_t(\theta, \gamma)\right\} \text{ is stable over } t \tag{5}$$

Usual GMM methods for estimation and inference use only the first k_Z full-sample restrictions $E\{(1/T)\sum_{t=1}^{T} Z_t' u_t(\theta, \gamma)\} = 0$. Magnusson and Mavroeidis (2010a) propose tests for the vector of structural parameters θ that explore both restrictions. Testing the assumption $H_0^{\theta}: \theta = \theta_0$ against $H_1^{\theta}: \theta \neq \theta_0$ can be indirectly conducted by testing both restrictions in (5), evaluated at θ_0, against the alternative that at least one of these conditions is violated. The tests have the following general form:

$$\text{gen-}S(\theta_0) = \text{gen-}\widetilde{S}(\theta_0; \widetilde{c}) + \frac{\overline{c}}{1+\overline{c}}\, S(\theta_0) \tag{6}$$

The first component of gen-S, the gen-$\widetilde{S}$, tests the stability restrictions, while its second component, the S test, detects violations of the full-sample moment restrictions. The nonnegative scalars $\widetilde{c}$ and $\overline{c}$ determine the weights that the investigator attaches to violations of the stability and full-sample restrictions under H_1, respectively. In this framework, the S test is considered a test that sets no weight on the stability restrictions; that is, $\widetilde{c} = 0$. Several possibilities exist for choosing gen-$\widetilde{S}$. Here we present four such statistics, which are described in the next section. The proposed stability tests are closely related to the quasi-local-level (qLL) test derived in Elliott and Müller (2006) and to the average (ave-), exponential (exp-), and supremum (sup-) Wald tests derived in Andrews (1993) and Sowell (1996). Additionally, all the proposed stability tests are asymptotically independent from the S test.

The four gen-S tests implemented by the `genstest` command are denoted qLL-S, ave-S, exp-S, and sup-S. In deriving the four tests, we set $\widetilde{c} = \overline{c}$; that is, violations of the full-sample and stability moment restrictions are weighted equally. In particular, in the case of qLL-S, $\widetilde{c} = \overline{c} = 10$. All the suggested tests have nontrivial power when instabilities are present under the alternative hypothesis. However, according to the weighted average power criteria, the qLL-S dominates the other tests if the instability of the moments follows a difference martingale sequence under H_1, and the ave-S and exp-S dominate the other tests if a single break is assumed at an unknown date.[2] Further details about the optimality properties of these tests are in Magnusson and Mavroeidis (2010a).

We can use the gen-S tests for estimating confidence intervals and sets. The $1 - \alpha$ confidence interval (set) consists of the points $\overline{\theta}$ in the parameter space Θ that do not reject the test under $H_0 : \theta = \overline{\theta}$ at α significance level. Once a grid of points in the parameter space is defined, we proceed by computing the tests at these points and selecting them accordingly.

The gen-S tests are asymptotically pivotal under H_0. Although their limit distributions are not standard, critical values can be simulated. Included with the `genstest` command are critical value tables up to the case where $k_Z = 20$ for all suggested tests.

3 The algorithm for implementing the generalized S tests

Next we show the algorithms for computing the two components of the gen-S test in (6), starting with the S test.

2. Set $\widetilde{c} = \overline{c} = c$. The ave-$S$ asymptotically power dominates the remaining tests when $c \longrightarrow 0$; the exp-S dominates when $c \longrightarrow +\infty$.

3.1 S test algorithm

The S test is obtained from the following steps:

1. Estimate the nuisance parameter vector γ under the null hypothesis $H_0^\theta : \theta = \theta_0$ using the following objective function,

$$\widehat{\gamma}(\theta_0) \equiv \arg\min_{\gamma} \; \mathbf{u}(\theta_0, \gamma)' \mathbf{Z} \left\{ \widehat{\Phi}(\theta_0) \right\}^{-1} \mathbf{Z}' \mathbf{u}(\theta_0, \gamma) \tag{7}$$

   where $\mathbf{u}(\theta_0, \gamma)$ is a $T \times 1$ vector whose typical tth element is $u_t(\theta_0, \gamma)$, $\mathbf{Z}$ is a $T \times k_Z$ matrix of instruments, and $\widehat{\Phi}(\theta_0)$ is an estimator of $\Phi(\theta_0, \gamma)$, the variance of $\mathbf{Z}'\mathbf{u}(\theta_0, \gamma)$.

2. Substitute $\mathbf{u}(\theta_0, \gamma)$ by $\mathbf{u}\{\theta_0, \widehat{\gamma}(\theta_0)\}$ into the objective function in (7).

3. The S test for testing $H_0^\theta : \theta = \theta_0$ is

$$S(\theta_0) = \mathbf{u}\{\theta_0, \widehat{\gamma}(\theta_0)\}' \mathbf{Z} \left\{ \widehat{\Phi}(\theta_0) \right\}^{-1} \mathbf{Z}' \mathbf{u}\{\theta_0, \widehat{\gamma}(\theta_0)\} \tag{8}$$

   Under the null hypothesis, $S(\theta_0) \xrightarrow{d} \chi^2_{(k_Z - q)}$, where $\chi^2_{(k_Z - q)}$ is a chi-squared distribution with $(k_Z - q)$ degrees of freedom.

The S test in (8) has two differences from the one proposed by Chernozhukov and Hansen (2008). First, the proposed S test encompasses models in which the residual term is a nonlinear function of the parameters (see examples 2 and 3 in section 5). Second, if the residual vector $\mathbf{u}(\theta_0, \gamma)$ is a linear function of parameters, then we concentrate the nuisance γ using an oblique projection matrix instead of a linear projection matrix.[3]

Next we turn to gen-$\widetilde{S}$, the stability part of the gen-S test.

3.2 The stability tests

There are two classes of tests that detect instabilities of the moments under the alternative assumption. The first class corresponds to the qLL-$\widetilde{S}$, a test calibrated to detect small but persistent changes in the moments. In the second class, the tests are derived assuming that there is only a single break in the moment at an unknown date. They are the ave-$\widetilde{S}$, the exp-$\widetilde{S}$, and the sup-$\widetilde{S}$.

3. When $\mathbf{u}(\theta_0, \gamma) = \mathbf{y} - \mathbf{Y}\theta_0 - \mathbf{X}\gamma$, substituting γ by its ordinary least-squares estimate $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(y - Y\theta_0)$ is the same as premultiplying $\mathbf{u}(\theta_0, \gamma)$ by $M_X = I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, the matrix that projects onto the orthogonal space spanned by the columns of $\mathbf{X}$. Our method is equivalent to premultiplying $\mathbf{u}(\theta_0, \gamma)$ by the oblique projection matrix $M_X^\Psi = I - \mathbf{X}(\mathbf{X}'\Psi\mathbf{X})^{-1}\mathbf{X}'\Psi$, where $\Psi = \mathbf{Z}\left\{\widehat{\Phi}(\theta_0)\right\}^{-1}\mathbf{Z}'$.

Persistent time-variation case

In the algorithm for computing the qLL-$\widetilde{S}$ test, we define the following $T \times T$ matrices and $T \times 1$ vector,

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ r & 1 & 0 & \cdots & \vdots \\ r^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & r & 1 & 0 \\ r^{T-1} & \cdots & r^2 & r & 1 \end{bmatrix}, \text{ and } \quad \mathbf{r} = \begin{bmatrix} r \\ r^2 \\ \vdots \\ \vdots \\ r^T \end{bmatrix}$$

where $r = 1 - (10/T)$. The $\mathbf{D}$ matrix is a first-difference operator, while $\mathbf{R}$ is the cumulative product operator matrix. Let $\widehat{\mathbf{U}}(\theta_0)$ be the following $T \times k$ matrix,

$$\widehat{\mathbf{U}}(\theta_0) = \{\widehat{\mathbf{u}}(\theta_0), \ldots, \widehat{\mathbf{u}}(\theta_0)\}$$

where $\widehat{\mathbf{u}}(\theta_0) = \mathbf{u}\{\theta_0, \widehat{\gamma}(\theta_0)\}$. The qLL-$\widetilde{S}$ statistic, which is the stability part of the qLL-S test, is obtained after taking the following steps:

1. First, compute the $T \times k$ matrix $\widehat{\mathbf{V}}(\theta_0) = \{\widehat{\mathbf{U}}(\theta_0) \odot \mathbf{Z}\}\widehat{\Phi}(\theta_0)^{-1/2}$, where $\odot$ denotes the direct product, and $\widehat{\Phi}(\theta_0)^{-1/2}$ is the symmetric square root matrix of $\widehat{\Phi}(\theta_0)^{-1}$. Second, compute $\widehat{\mathbf{H}}(\theta_0)=\mathbf{R}\{\mathbf{D}\widehat{\mathbf{V}}(\theta_0)\}$, a $T \times k$ matrix.

2. Estimate the $T \times k$ matrix $\widehat{\mathbf{w}}$, the ordinary least-square residuals of the following regression,
$$\widehat{\mathbf{H}}(\theta_0) = \mathbf{r}B + \mathbf{w}$$
where B is a $1 \times k$ row vector of parameters, and compute $\text{TSSR}_{\widehat{\mathbf{w}}} = \sum_{i=1}^{k}\sum_{t=1}^{T}(\widehat{w}_{i,t})^2$, the total sum of squared residuals of the above regression.

3. Compute the $T \times k$ matrix $\widehat{\boldsymbol{\epsilon}}$, the ordinary least-square residuals of the following regression,
$$\widehat{\mathbf{V}}(\theta_0) = \boldsymbol{\iota}_T C + \boldsymbol{\epsilon}$$
where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of 1s, and C is a $1 \times k$ row vector of parameters. Calculate $\text{TSSR}_{\widehat{\boldsymbol{\epsilon}}} = \sum_{i=1}^{k}\sum_{t=1}^{T}(\widehat{\epsilon}_{i,t})^2$, the total sum of squared residuals of the above regression.

4. The qLL-$\widetilde{S}$ statistic under H_0 is
$$\text{qLL-}\widetilde{S}(\theta_0) = \text{TSSR}_{\widehat{\boldsymbol{\epsilon}}} - r \times \text{TSSR}_{\widehat{\mathbf{w}}}$$
The asymptotic distribution of this statistic is a functional of a k-dimensional Ornstein–Uhlenbeck process.

5. The qLL-S test is defined as
$$\text{qLL-}S(\theta_0) = \text{qLL-}\widetilde{S}(\theta_0) + \frac{10}{11}S(\theta_0)$$

Single break, unknown break date

We take the following steps in computing the stability part of the ave-, exp-, and sup-$\widetilde{S}$ tests:

1. Specify an interval where a break in the moments is located. This interval must be defined as $[t_l, t_u]$, where $t_l = [sT]$, $t_u = [(1-s)T]$, $s \in (0, 0.5)$, and $[m]$ denotes the integer part of m.
2. For a possible break date $j \in [t_l, t_u]$, let $T_1 = j$, and $T_2 = T - j$. Partition $\mathbf{Z}$ as $\mathbf{Z} = (\mathbf{Z}_1' : \mathbf{Z}_2')'$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are $T_1 \times k$ and $T_2 \times k$ submatrices of $\mathbf{Z}$ containing, respectively, observations before and after j. Similarly, partition $\mathbf{u}(\theta_0, \gamma)$ as $\mathbf{u}(\theta_0, \gamma) = \left\{\mathbf{u}_1(\theta_0, \gamma)' : \mathbf{u}_2(\theta_0, \gamma)'\right\}'$.
3. Estimate the nuisance parameter γ under the null hypothesis $H_0 : \theta = \theta_0$. Similar to the S test, this step consists of solving the following minimization problem,
$$\begin{aligned}\widehat{\gamma}_j(\theta_0) \equiv \arg\min_{\gamma} \; & \mathbf{u}_1(\theta_0, \gamma)' \mathbf{Z}_1 \left\{\widehat{\Phi}_1(\theta_0)\right\}^{-1} \mathbf{Z}_1' \mathbf{u}_1(\theta_0, \gamma) \\ & + \mathbf{u}_2(\theta_0, \gamma)' \mathbf{Z}_2 \left\{\widehat{\Phi}_2(\theta_0)\right\}^{-1} \mathbf{Z}_2' \mathbf{u}_2(\theta_0, \gamma) \end{aligned} \tag{9}$$
where $\widehat{\Phi}_1(\theta_0)$ and $\widehat{\Phi}_2(\theta_0)$ are, respectively, estimators of $\Phi_1(\theta_0, \gamma)$ and $\Phi_2(\theta_0, \gamma)$, the variances of $\mathbf{Z}_1'\mathbf{u}_1(\theta_0, \gamma)$ and $\mathbf{Z}_2'\mathbf{u}_2(\theta_0, \gamma)$ under H_0.
4. Substitute $\mathbf{u}_1(\theta_0, \gamma)$ and $\mathbf{u}_2(\theta_0, \gamma)$ with $\mathbf{u}_1\{\theta_0, \widehat{\gamma}_j(\theta_0)\}$ and $\mathbf{u}_2\{\theta_0, \widehat{\gamma}_j(\theta_0)\}$, respectively, into the objective function in (9).
5. Compute the following modified S test assuming a break at date j:
$$\begin{aligned} S(\theta_0; j) = \; & \mathbf{u}_1\{\theta_0, \widehat{\gamma}_j(\theta_0)\}' \mathbf{Z}_1 \left\{\widehat{\Phi}_1(\theta_0)\right\}^{-1} \mathbf{Z}_1' \mathbf{u}_1\{\theta_0, \widehat{\gamma}_j(\theta_0)\} \\ & + \mathbf{u}_2\{\theta_0, \widehat{\gamma}_j(\theta_0)\}' \mathbf{Z}_2 \left\{\widehat{\Phi}_2(\theta_0)\right\}^{-1} \mathbf{Z}_2' \mathbf{u}_2\{\theta_0, \widehat{\gamma}_j(\theta_0)\} \end{aligned} \tag{10}$$
Define the following statistic:
$$\widetilde{S}(\theta_0; j) = S(\theta_0; j) - S(\theta_0)$$
6. Repeat steps 1 through 5 for each possible break date in $[t_l, t_u]$.
7. The ave-, exp-, and sup-$\widetilde{S}$ tests are defined as
$$\begin{aligned} \text{ave-}\widetilde{S}(\theta_0) &= \frac{1}{d(t_l, t_u)} \sum_{j=t_l}^{t_u} \widetilde{S}(\theta_0; j) \\ \text{exp-}\widetilde{S}(\theta_0) &= 2\log\left[\frac{1}{d(t_l, t_u)} \sum_{j=t_l}^{t_u} \exp\left\{\frac{1}{2}\widetilde{S}(\theta_0; j)\right\}\right] \\ \text{sup-}\widetilde{S}(\theta_0) &= \sup_{j \in [t_l, t_u]} \widetilde{S}(\theta_0; j) \end{aligned}$$

where $d(t_l, t_u) = t_u - t_l + 1$. One can show that the asymptotic distributions of the above tests are functionals of standard k-dimensional Brownian bridge processes on $(0,1)$.

8. The ave-, exp-, and sup-S tests are defined, respectively, as

$$\begin{aligned}\text{ave-}S(\theta_0) &= S(\theta_0) + \text{ave-}\widetilde{S}(\theta_0)\\ \text{exp-}S(\theta_0) &= S(\theta_0) + \text{exp-}\widetilde{S}(\theta_0)\\ \text{sup-}S(\theta_0) &= S(\theta_0) + \text{sup-}\widetilde{S}(\theta_0)\end{aligned}$$

3.3 The estimation of nuisance parameters and variance–covariance matrix

The estimators of $\widehat{\gamma}(\theta_0)$ and $\widehat{\gamma}_j(\theta_0)$ in equations (7) and (9) can be the two-step or iterative GMM estimators. The first-step estimator for computing $\widehat{\gamma}(\theta_0)$, necessary for estimating Φ, solves

$$\min_{\gamma} \; \mathbf{u}(\theta_0, \gamma)' \mathbf{Z} \, \mathcal{W} \, \mathbf{Z}' \mathbf{u}(\theta_0, \gamma) \tag{11}$$

where $\mathcal{W}$ is a square matrix (for example, the identity matrix or $(\mathbf{Z}'\mathbf{Z})^{-1}$). Similarly, the first step for computing $\widehat{\gamma}_j(\theta_0)$ solves

$$\min_{\gamma} \; \mathbf{u}_1(\theta_0, \gamma)' \mathbf{Z}_1 \mathcal{W}_1 \mathbf{Z}_1' \mathbf{u}_1(\theta_0, \gamma) + \mathbf{u}_2(\theta_0, \gamma)' \mathbf{Z}_2 \mathcal{W}_2 \mathbf{Z}_2' \mathbf{u}_2(\theta_0, \gamma)$$

where $\mathcal{W}_1$ and $\mathcal{W}_2$ are conformable quadratic matrices. For theoretical reasons, $\widehat{\gamma}(\theta_0)$ and $\widehat{\gamma}_j(\theta_0)$ cannot be the first-step estimators of γ; see Stock and Wright (2000) and Caner (2007). Under the null assumption and under sequences of local alternatives, $\widehat{\gamma}(\theta_0)$ and $\widehat{\gamma}_j(\theta_0)$ have the same probability limits.[4] Hence, we can replace $\widehat{\gamma}_j(\theta_0)$ with $\widehat{\gamma}(\theta_0)$ for computing the $S(\theta_0; j)$ statistic in (10).

The estimation of the variance–covariance matrix Φ depends on the assumption about the asymptotic variance of $T^{-1/2}\sum_{t=1}^{T} Z_t' u_t(\theta, \gamma)$ evaluated at $\{\theta_0, \widehat{\gamma}(\theta_0)\}$. The `genstest` command is very flexible about the structure of the variance matrix, allowing for homoskedastic residuals; heteroskedastic residuals (including adjustment factors `hc1`, `hc2`, `hc3`, and `hc4`; see Davidson and MacKinnon [2003]); cluster residuals; and heteroskedastic autocorrelated residuals (including options for the kernel and number of lags when computing the autocorrelation terms). More details are in the following section.

The general form of the estimator $\widehat{\Phi}(\theta_0)$ in (7) used for computing the S and qLL-S tests is $\mathbf{Z}'\widehat{\boldsymbol{\Omega}}(\theta_0)\mathbf{Z}$, where $\widehat{\boldsymbol{\Omega}}(\theta_0)$ is a $T \times T$ matrix whose elements are a function of the vector of the estimated residuals $\mathbf{u}\{\theta_0, \widehat{\gamma}(\theta_0)\}$.

In the case of ave-, exp-, and sup-S, the estimators $\widehat{\Phi}_1(\theta_0)$ and $\widehat{\Phi}_2(\theta_0)$ in (10) can be represented by the $T_i \times T_i$ matrix $\mathbf{Z}_i \widehat{\boldsymbol{\Omega}}_i(\theta_0)\mathbf{Z}_i$ for $i = 1, 2$. The elements of $\widehat{\boldsymbol{\Omega}}_i(\theta_0)$ are

4. However, they have different limits under a fixed sequence of alternatives; see Andrews (1993).

a function of $\mathbf{u}_i\{\theta_0, \widehat{\gamma}_j(\theta_0)\}$. The asymptotic variances of $T_1^{-1/2}\mathbf{Z}_1'\mathbf{u}_1\{\theta_0, \widehat{\gamma}_j(\theta_0)\}$ and $T_2^{-1/2}\mathbf{Z}_2'\mathbf{u}_2\{\theta_0, \widehat{\gamma}_j(\theta_0)\}$ are the same as the asymptotic variance of $T^{-1/2}\mathbf{Z}'\mathbf{u}\{\theta_0, \widehat{\gamma}(\theta_0)\}$ (see Hall [2005]), so we can substitute $\widehat{\Phi}_1(\theta_0)$ and $\widehat{\Phi}_2(\theta_0)$ with $\widehat{\Phi}(\theta_0)$ in (10).

4 The genstest command

The `genstest` command implements the above four gen-S tests in Stata and Mata. It may be invoked as a stand-alone command or as a postestimation command for `gmm`.

When `genstest` is used as a postestimation command, it will only use the `gmm` options that `genstest` implements (they are listed and described in this section). Any additional `gmm` options will be discarded. Furthermore, `genstest` performs the tests on only one residual expression, so the `gmm` estimation command should conform to that limitation.

Additionally, the command can estimate confidence intervals and sets (up to two parameters) based on these tests.[5] These intervals (sets) are generated using a grid search method on the points of the parameter space that do not reject the null hypothesis of the test.

The `genstest` command requires at least Stata 10 because of the use of Mata's optimization functions. No additional packages are required beyond a standard Stata and Mata installation.

4.1 Syntax

The syntax for `genstest` was designed to be as similar to Stata 11's `gmm` command as possible. The syntax is defined as follows:

`genstest` [`(`*residual*`)`] [*if*] [*weight*] [`,` `instruments(`*varlist* [`, noconstant`]`)`
`derivative(/`*name* `=` [`<`] *dexp* [`>`]`)` `twostep` `igmm` `init(`*numlist*`)`
`null(`*numlist*|`last)` `test(`*namelist*`)` `sb` `stab` `winitial(`*iwtype*`)`
`wmatrix(`*wmtype*`)` `center` `small` `trim(`*#*`)` `nuisS` `varS` `ci(`*ci_options*`)`]

residual is an expression defining the $u_t(\theta, \gamma)$, the error-term function used in the empirical moment $(1/T)\sum_{t=1}^{T} Z_t' u_t(\theta_0, \gamma)$, where Z_t is the vector of instruments.

In the *residual* expression, enclosing a name inside `<>` indicates a parameter to be tested as the null hypothesis, while enclosing a name inside `{}` indicates a parameter to estimate. For example, in the following linear regression model,

$$y_{1,t} = y_{2,t}\theta + x_t\gamma + u_t$$

5. Because `genstest` can perform hypothesis tests on any number of parameters, to generate a confidence set for a higher number of parameters, one only needs to use nested `while` loops.

where θ is the parameter to be tested, and γ is estimated under the null hypothesis. The regression residual expression is

```
(y1 - <theta>*y2 - {gamma}*x)
```

A constant is not added to the residual expression by default to keep the behavior of genstest similar to that of the gmm command. However, a constant is automatically included in the vector of instruments, Z_t, unless noconstant is specified in the instruments() option.

In the same example, if the residual expression is specified as

```
(y1 - {theta}*y2 - {gamma}*x)
```

then both parameters are estimated. In this case, the S test will be the same as the overidentification restriction of Hansen (1982), also known as the J test, and the ave-, exp-, and sup-S tests will be equivalent to the overidentification restriction tests proposed by Hall and Sen (1999), also known as the O test.

When running genstest as a postestimation command, one does not need to specify *residual*. In that case, genstest uses the *residual* given to gmm.

4.2 Testing options

instruments(*varlist*[, noconstant]) specifies the vector of instruments Z_t. The optional noconstant indicates removal of a constant from the matrix of instruments.

derivative(/*name* = [<]*dexp*[>]) specifies the derivative of a *residual* function with respect to the parameter *name*. The functionality of this option requires entering all untested parameter derivatives; otherwise, derivatives in the optimization algorithm will be computed numerically. The use of this option is recommended when estimating confidence intervals and sets because it improves the performance of the optimization algorithm and the computational speed (see ci() option). This option is specified as in gmm with the addition that the <>'s indicate the value of a parameter tested under the null hypothesis. If one is using genstest as a postestimation command, the derivatives passed to gmm will be used by genstest.

twostep requires the two-step general method of moments estimator is used (this is the default).

igmm requires the iterated general method of moments estimator be used.

init(*numlist*) sets the initial values in the optimization routine for estimating the nuisance parameters. The default choice is a vector of zeros. One should include this option if the algorithm for estimating untested parameters does not converge, if it converges to a local minimum, or if the residual expression is undefined at the zero vector.

`null(`*numlist* | `last)` tells `genstest` to test $H_0^\theta : \theta = \theta_0$, where θ_0 is the hypothesized value of the parameter of interest θ. By default, `genstest` tests the null hypothesis that all parameters defined inside `<>` are equal to 0. If `genstest` is used as a postestimation command, then `null(last)` will set θ_0 as the last estimate obtained after running `gmm`. The supplied *numlist* must be in the same order as the parameters appearing in the residual expression.

`test(`*namelist*`)` lists the names of the `gmm` parameters in the *residual* expression to be tested. This option is only applicable if `genstest` is being used as a postestimation command.

`sb` reports the ave-, exp-, and sup-S tests (the single-break tests). These tests are computationally more intensive than the qLL-S test and therefore not computed by default.

`stab` reports the stability tests (the $\widetilde{S}$ tests).

`winitial(`*iwtype*`)` specifies the initial weighting matrix $\mathcal{W}$ in (11) for obtaining an inefficient estimate of γ. There are two options for this matrix: `identity`, which uses the identity matrix, and `unadjusted`, which sets $(\mathbf{Z}'\mathbf{Z})$ as the initial weight matrix. The default is `winitial(unadjusted)`.

`wmatrix(`*wmtype*`)` allows the choice of the covariance matrix in (7). *wmtype* represents the user choice for the estimator type of the variance of $T^{-1/2}\sum_{t=1}^{T} Z_t' u_t(\theta_0, \gamma)$. The choices are the following:

`unajusted` for the homoskedastic case.

`robust`, `hc1`, `hc2`, `hc3`, and `hc4` for the heteroskedastic case with `hc`*i*, for $i = 1, \ldots, 4$ denoting the residual adjustment options (see Davidson and MacKinnon [2003]). The default is `wmatrix(hc1)`, which denotes multiplying the square of the residuals by $\{T/(T - k_Z)\}$.

`cluster` *clustvar* for a cluster–robust covariance matrix having the cluster variable defined in *clustvar*.

`hac` *kernel*[*lags*] for the heteroskedastic and autocorrelated (HAC) robust covariance matrix. The *kernel* can be defined as

`bartlett` or `nwest` for the Bartlett (Newey–West) kernel;

`parzen` or `gallant` for the Parzen (Gallant) kernel; or

`quadraticspectral` or `andrews` for the quadratic spectral (Andrews) kernel.

When selecting the kernel, the user can choose the number of *lags* to compute the HAC estimate. *lags* may be one of the following:

`optimal`, if using the optimal selection algorithm of Newey and West (1994) (implemented using the same algorithm as in `gmm`).

`automatic` for setting the number of lags to the starting value of the optimal-lag selection algorithm (divided by 5):

Bartlett: $4 \times \left(\frac{T}{100}\right)^{\frac{2}{9}}$

Parzen: $4 \times \left(\frac{T}{100}\right)^{\frac{4}{25}}$

Quadratic Spectral: $4 \times \left(\frac{T}{100}\right)^{\frac{2}{25}}$

number for any number specified by the user.

❏ **Technical note**

The `genstest` default number of lags is `automatic`, which differs from the default number in the built-in Stata `gmm` function.

❏

`center` indicates recentering the moment function when computing the HAC estimate of the variance.

`small` indicates using a small-sample adjustment when computing the HAC weight matrix.

Options: Single-break tests

The following options are for computation of the ave-, exp-, and sup-S tests.

`trim(`*#*`)` specifies the value of the trimming parameter s used to fix $t_l = [sT]$ and $t_u = [(1-s)\,T]$ in step 1 of the algorithm for computing the single-break stability tests. The options are $s = 0.05$, 0.10, 0.15, and 0.20. The default is `trim(0.15)`.

`nuisS` indicates the use of $\widehat{\gamma}\,(\theta_0)$, the estimate of the nuisance parameter in (7), in place of $\widehat{\gamma}_j\,(\theta_0)$, derived from (9), when computing the split-sample tests.

`varS` specifies the use of $\widehat{\Phi}\,(\theta_0)$, the estimated variance of the moments using all observations, in place of both $\widehat{\Phi}_1\,(\theta_0)$ and $\widehat{\Phi}_2\,(\theta_0)$ when computing the split-sample tests.

The weight matrices for computing $\widehat{\Phi}_1\,(\theta_0)$ and $\widehat{\Phi}_2\,(\theta_0)$ are the same as the one for `winitial()` and `wmatrix()`. For example, choosing `winitial(unadjusted)` and `wmatrix(hac nwest automatic)` implies that the initial weight matrices for estimating $\widehat{\gamma}(\theta_0; j)$ are $(\mathbf{Z}_1'\mathbf{Z}_1)$ and $(\mathbf{Z}_2'\mathbf{Z}_2)$ in (3.3) and that the HAC estimators use the Bartlett kernel with lags $4 \times (T_1/100)^{2/9}$ and $4 \times (T_2/100)^{2/9}$ for $\widehat{\Phi}_1\,(\theta_0)$ and $\widehat{\Phi}_2\,(\theta_0)$, respectively.

Options: Confidence interval and region

The `genstest` command has the option of estimating confidence intervals and sets, the latter up to two parameters, using a grid search algorithm (to estimate higher-dimensional sets, one can use a simple `while` loop because `genstest` can test any number of parameters). If a two-parameter confidence set is chosen, the result can be displayed in a `twoway` graph. The use of the option `derivative()` is recommended when estimating confidence intervals and sets. The options are the following:

`ci(`*numlist* [`,` *ci_options*]`)` indicates that a confidence interval or set be estimated.

numlist specifies the range of the grid search. For example,

`ci(a b c d,` *ci_options*`)`

sets [a,b] and [c,d] as the grid search range of a confidence region for two parameters.

`points(`*numlist*`)` determines the number of equally spaced points for the grid search. The default is `points(20)` for confidence intervals and `points(20 20)` for confidence sets.

`alpha(`*#*`)` determines the $1-\alpha$ coverage probability of the interval or set. The default is `alpha(0.05)`.

`allpv` tells `genstest` to return *p*-values for all points tested in the selected range. Therefore, if one wishes to examine the confidence interval (set) for a different significance level, there is no need to execute the command a second time.

`autograph` tells `genstest` to automatically graph the confidence region if two parameters are being tested. Whether or not this option is specified, the points necessary to plot the confidence region are stored in matrices.

❏ Technical note

The estimation of confidence intervals and sets, which is based on a grid search process, is not in the default of `genstest`, because it can be computationally intensive. To estimate confidence intervals and regions, we recommend running without the `sb` option (which would cause `genstest` to perform the split-sample tests) or using the `sb` option with the `nuisS` and `varS` options.

❏

❏ Technical note

For confidence interval results, if `allpv` is not given, then `genstest` returns a matrix of values that pass the test alongside the resulting statistic. If `allpv` is specified, `genstest` saves a matrix containing the grid search values associated with their respective *p*-values (*p*-values and grid search points are reported in the first and the subsequent columns, respectively).

❏

4.3 Stored results

`genstest` stores the following in `r()`:

Scalars

`r(S)`	S statistic
`r(aveS)`	ave-S statistic
`r(expS)`	exp-S statistic
`r(supS)`	sup-S statistic
`r(qllS)`	qLL-S statistic
`r(avestabS)`	ave-$\widetilde{S}$ statistic
`r(expstabS)`	exp-$\widetilde{S}$ statistic
`r(supstabS)`	sup-$\widetilde{S}$ statistic
`r(qllstabS)`	qLL-$\widetilde{S}$ statistic
`r(pS)`	S statistic p-value
`r(paveS)`	ave-S p-value
`r(pexpS)`	exp-S p-value
`r(psupS)`	sup-S statistic p-value
`r(pqllS)`	qLL-S p-value
`r(pavestabS)`	ave-$\widetilde{S}$ p-value
`r(pexpstabS)`	exp-$\widetilde{S}$ p-value
`r(psupstabS)`	sup-$\widetilde{S}$ p-value
`r(pqllstabS)`	qLL-$\widetilde{S}$ p-value

Matrices

`r(Sci)`	grid search points not rejected by the S test or search points and their associated p-values (if `allpv` is specified)
`r(aveSci)`	grid search points not rejected by the ave-S test or search points and their associated p-values
`r(expSci)`	grid search points not rejected by the exp-S test or search points and their associated p-values
`r(supSci)`	grid search points not rejected by the sup-S test or search points and their associated p-values
`r(qllSci)`	grid search points not rejected by the qLL-S test or search points and their associated p-values
`r(avestabSci)`	grid search points not rejected by the ave-$\widetilde{S}$ test or search points and their associated p-values
`r(expstabSci)`	grid search points not rejected by the exp-$\widetilde{S}$ test or search points and their associated p-values
`r(supstabSci)`	grid search points not rejected by the sup-$\widetilde{S}$ test or search points and their associated p-values
`r(qllstabSci)`	grid search points not rejected by the qLL-$\widetilde{S}$ test or search points and their associated p-values

5 Examples

We present three examples to illustrate the use of `genstest`. The first example is the regression model of married female labor supply presented in Mroz (1987). The second example is based on the Poisson regression model contained in the example section of the `gmm` command in Stata's *Base Reference Manual: Release 11* (see StataCorp [2009]). In the first two examples, we assume that the observations are independent but not identically distributed. The third example presents inference about parameters of the new Keynesian Phillips curve (NKPC) model discussed in Sbordone (2005) and

Magnusson and Mavroeidis (2010b). In this last example, residuals are assumed to be heteroskedastic and exhibit autocorrelation of unknown form.

5.1 Example 1: Instrumental variable regression—independent observations and heteroskedastic residuals

In studying the married female labor supply, Mroz (1987) suggests a regression of hours of work (`hours`) on the log of wages (`lwage`, the only endogenous variable), household income excluding the woman's wage (`nwifeinc`), years of education (`educ`), age (`age`), and the number of children less than six and greater than six years old (`kidslt6` and `kidsge6`, respectively). The chosen excluded instruments are the actual labor market experience and its square (`exper` and `expersq`) and the father's and mother's years of education (`fatheduc` and `motheduc`). The data consist of 428 women in the labor force. The IV regression model may be summarized as follows:

$$\begin{aligned}\texttt{hours} &= \theta\,\texttt{lwage} + \gamma_0 + \gamma_1\texttt{nwifeinc} + \gamma_2\texttt{educ} + \gamma_3\texttt{age} + \gamma_4\texttt{kidslt6} + \gamma_5\texttt{kidsge6} + u \\ \texttt{lwage} &= \rho_0 + \rho_1\texttt{exper} + \rho_2\texttt{expersq} + \rho_3\texttt{fatheduc} + \rho_4\texttt{motheduc} \\ &\quad + \rho_5\texttt{educ} + \rho_6\texttt{nwifeinc} + \rho_7\texttt{age} + \rho_8\texttt{kidslt6} + \rho_9\texttt{kidsge6} + v\end{aligned}$$

We examine the effect of `lwage` on hours of work. We sort the data by `lwage` because one might be concerned if this effect is constant across observations. Sorting the data does not affect Wald and S tests. Independence among observations is assumed, but the distribution of the error term is heteroskedastic. Because `genstest` uses a weight matrix robust to heteroskedasticity by default (`hc1`), the option is omitted below.

```
. use http://www.stata.com/data/jwooldridge/eacsap/mroz.dta
. sort lwage
. gmm (hours - {theta}*lwage - {g0} - {g1}*educ - {g2}*nwifeinc - {g3}*age -
> {g4}*kidslt6 - {g5}*kidsge6) if inlf==1, level(90)
> inst(exper expersq fatheduc motheduc educ nwifeinc age kidslt6 kidsge6)
  (output omitted)
GMM estimation
Number of parameters =   7
Number of moments    =  10
Initial weight matrix: Unadjusted                    Number of obs  =     428
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   1223.656   456.8492     2.68   0.007      472.206    1975.106
         /g0 |   2287.937   522.7613     4.38   0.000     1428.071    3147.803
         /g1 |  -143.8503   52.84411    -2.72   0.006    -230.7712    -56.9295
         /g2 |  -8.466459   4.488806    -1.89   0.059    -15.84989   -1.083031
         /g3 |  -8.105428   8.896522    -0.91   0.362     -22.7389    6.528048
         /g4 |  -261.7084    177.988    -1.47   0.141    -554.4726    31.05575
         /g5 |  -56.63245   48.20149    -1.17   0.240    -135.9168    22.65194
------------------------------------------------------------------------------
Instruments for equation 1: exper expersq fatheduc motheduc educ nwifeinc
    age kidslt6 kidsge6 _cons
```

```
. genstest (hours - <theta>*lwage - {g0} - {g1}*educ - {g2}*nwifeinc - {g3}*age -
> {g4}*kidslt6 - {g5}*kidsge6) if inlf==1,
> inst(exper expersq fatheduc motheduc educ nwifeinc age kidslt6 kidsge6)
> stab wmatrix(hc1) ci(-200 7000, points(60) alpha(0.10))
```

Test	Statistic	P-value	CI (alpha=.1)
S	26.316010	0.000	[880, 6280]
qLL-S	68.829101	0.006	Rejected Grid
qLL-stab-S	42.513092	0.632	[-80, 280]

```
Tested null hypothesis vector: <theta> = < 0.000>
Number of Instruments - Included Instruments: 10 - 6
Number of Observations: 428
```

The difference between the Wald and S confidence intervals for θ indicates the presence of weak instruments. It is interesting to observe that the S test confidence interval does not intersect the qLL-$\widetilde{S}$ confidence interval. This results in an empty qLL-S confidence interval and clearly suggests that the effect of `lwage` on hours of work is not constant.

This example uses `genstest` as a stand-alone command. Because `genstest` uses the null that all parameters of interest are equal to 0 by default, the option `null()` is omitted. In such cases, there is no need to run `gmm` first (we do so to show the Wald confidence interval for comparison).

5.2 Example 2: Exponential regression with endogenous regressors

This example corresponds to [R] **gmm** examples 6, 7, and 8 on pages 591–595 in Stata's *Base Reference Manual* (see StataCorp [2009]). Cameron and Trivedi (2010) model doctor visits on the basis of the following factors: a patient's income (`income`), whether a patient has a chronic disease (`chronic`), whether a patient has private insurance (`private`), and gender. They use an exponential regression model. The dataset has demographic information on 4,412 patients. Taking income to be endogenous, one adds these additional instruments: `age` and the dummy variables `hispanic` and `black`. We subset the model to include only female patients (there are 2,082 observations). The components of the empirical moment are

$$u = \texttt{docvis} - exp(\theta\texttt{income} + \gamma_0 + \gamma_1\texttt{chronic} + \gamma_2\texttt{private})$$
$$Z = (1,\ \texttt{chronic},\ \texttt{private},\ \texttt{age},\ \texttt{black},\ \texttt{hispanic})$$

This Poisson regression model assumes that the residuals are heteroskedastic. Before performing inference using the `genstest` command, we sort the data according to `income` first and then `age` because of the possibility that lower- and higher-income groups have different income effects. We first estimate the parameters using the `gmm` command and then run the `genstest` as a postestimation command.

```
. webuse docvisits
. sort income age
. local expr = "exp({theta}*income + {g0} + {g1}*chronic + {g2}*private)"
. gmm (docvis - `expr´) if female==1, inst(age private chronic black hispanic)
> deriv(/theta = -1*income*`expr´) deriv(/g0 = -1*`expr´)
> deriv(/g1 = -1*chronic*`expr´) deriv(/g2 = -1*private*`expr´) level(99)
  (output omitted)
GMM estimation

Number of parameters =   4
Number of moments    =   6
Initial weight matrix: Unadjusted                     Number of obs  =    2082
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[99% Conf. Interval]	
/theta	.0178061	.0038447	4.63	0.000	.0079029	.0277094
/g0	.0791596	.1278643	0.62	0.536	-.250197	.4085162
/g1	.9640517	.076972	12.52	0.000	.7657849	1.162319
/g2	.4172219	.1590262	2.62	0.009	.0075976	.8268461

```
Instruments for equation 1: age private chronic black hispanic _cons
. genstest, null(last) test(theta) stab sb varS nuisS ci(-0.05 0.05, points(20)
> alpha(0.01))
convergence not achieved
```

Test	Statistic	P-value	CI (alpha=.01)
S	3.643981	0.303	[.01, .03]
qLL-S	2.02e+02	0.001	[.005, .005]
ave-S	22.440894	0.003	[.005, .015]
exp-S	29.458495	0.001	[.005, .01]
sup-S	35.183634	0.002	[0, .015]
qLL-stab-S	1.98e+02	0.001	[0, .005]
ave-stab-S	18.796914	0.001	[-.01, .01]
exp-stab-S	25.814514	0.001	[-.005, .01]
sup-stab-S	31.539654	0.001	[-.01, .015]

```
Tested null hypothesis vector: <theta> = < 0.018  >
Number of Instruments - Included Instruments: 6 - 3
Number of Observations: 2082
```

The `null(last)` option has `genstest` test the null hypothesis that θ is equal to its `gmm` estimated value. Here it is not necessary to specify the `derivative()` option in `genstest` again, because the command will use the `derivative()` expression of `gmm` and automatically consider `theta` as the only parameter to be tested. This command could also have been executed in the following way (assuming it is still being used as a postestimation command):

```
. local expr = "exp(<theta>*income + {g0} + {g1}*chronic+ {g2}*private)"
. genstest, null(last) test(theta) deriv(/g0 =-1*`expr´)
> deriv(/g1 = -1*chronic*`expr´) deriv(/g2 = -1*private*`expr´)
```

We use the options `varS` and `nuisS` to reduce computation time of the ave-, exp-, and sup-S tests. The p-values of the gen-S and gen-$\widetilde{S}$ tests indicate that at the 1%

significance level, we reject $H_0 : \theta = \widehat{\theta}$, where $\widehat{\theta}$ is the `gmm` estimate of θ. We also notice that the gen-S tests for 99% confidence interval lengths for θ are smaller than the length of the confidence interval reported by `gmm`. The reduction is due to imposing stability restrictions as indicated by the upper bounds of gen-$\widetilde{S}$ confidence intervals.

5.3 Example 3: NKPC

The hybrid NKPC is defined by the following equation,

$$\pi_t = \gamma + \frac{1}{1+\rho} E_t(\pi_{t+1}) + \frac{\rho}{1+\rho}\pi_{t-1} + \frac{(1-\phi)^2}{\phi(1+\rho)} x_t + \epsilon_t$$

where π_t is inflation, x_t is labor share, and ϵ_t is a shock. The parameter ρ measures the degree of indexation to past inflation. The parameter ϕ is the probability that a firm will be unable to change its price in a given period (hence, $1/(1-\phi)$ is the average time over which a price is fixed).

The empirical moment condition derived from the above model is

$$\underbrace{Z_t' \left\{ \Delta\pi_t - \gamma - \frac{1}{1+\rho}(\pi_{t+1} - \pi_{t-1}) - \frac{(1-\phi)^2}{\phi(1+\rho)} x_t \right\}}_{u_t(\theta,\gamma)}$$

where $Z_t = (1, \Delta\pi_{t-1}, \Delta\pi_{t-2}, x_{t-1}, x_{t-2}, x_{t-3})$, and $\theta = (\rho, \phi)$.

We illustrate first the estimation of confidence intervals for ϕ and ρ on the basis of the generalized S and $\widetilde{S}$ tests. We use quarterly data on inflation and labor share. Inflation is calculated from the gross domestic product deflator, while labor share is obtained from the Bureau of Labor Statistics and transformed according to the procedure used in Sbordone (2005). The data comprise information from 1959:2 to 2008:3.

The `genstest` command is used as a postestimation command. In estimating the confidence intervals, we restrict the grid search to be between 0 and 1, which corresponds to the range determined by the economic theory. All the options in `genstest` are set according to the `gmm` command option: a HAC weight matrix with the Bartlett kernel and the use of recentered moments for computing the HAC with the number of lags selected according to the optimal method suggested by Newey and West (1994) (see [R] **gmm** in Stata 11).[6]

6. We set 1 as the initial value for estimating ϕ. The default value for estimating ϕ using `gmm` is 0, which will result in an error message.

```
. use nkpc_gmm

. local expr = "{g} - (1/(1 + {rho}))*(F.inf - L.inf) -
> (((1 - {phi=1})^2)/({phi=1}*(1 + {rho})))*ls"

. generate time=q(1947q2)+_n-1 // generate a quarterly series

. format time %tq

. tsset time
        time variable:  time, 1947q2 to 2008q3
                delta:  1 quarter

. generate dinf = inf - L.inf
(1 missing value generated)

. gmm (dinf - `expr´) if time>=tq(1959q2) & time<=tq(2008q3),
> inst(L.dinf L2.dinf L.ls L2.ls L3.ls) wmat(hac nwest optimal) center
warning: 1 missing value returned for equation 1 at initial values
```

(*output omitted*)

```
GMM estimation

Number of parameters =   3
Number of moments    =   6
Initial weight matrix: Unadjusted                    Number of obs  =     197
GMM weight matrix:     HAC Bartlett 23
                       (lags chosen by Newey-West)
```

	Coef.	HAC Std. Err.	z	P>|z|	[95% Conf.	Interval]
/g	-.0045429	.0049064	-0.93	0.354	-.0141594	.0050735
/rho	.3862696	.1429788	2.70	0.007	.1060362	.6665029
/phi	.8208458	.0739276	11.10	0.000	.6759503	.9657412

```
HAC standard errors based on Bartlett kernel with 23 lags.
   (Lags chosen by Newey-West method.)
Instruments for equation 1: L.dinf L2.dinf L.ls L2.ls L3.ls _cons

. genstest, init(0 0.76) test(rho) null(last) ci(0.01 0.99, points(20)) stab sb
Note: using the nuisS and/or varS options will decrease computation time for
> the single-break tests.
```

Test	Statistic	P-value	CI (alpha=.05)
S	3.020303	0.554	[.059, .794]
qLL-S	35.709204	0.173	[.108, .696]
ave-S	12.124914	0.248	[.157, .549]
exp-S	16.133502	0.150	[.255, .5]
sup-S	20.743887	0.192	[.206, .5]
qLL-stab-S	32.688901	0.123	[.01, .941]
ave-stab-S	9.104611	0.101	[.157, .451]
exp-stab-S	13.113198	0.068	[.304, .451]
sup-stab-S	17.723584	0.113	[.255, .451]

```
Tested null hypothesis vector: <rho> = < 0.386  >
Number of Instruments - Included Instruments: 6 - 2
Number of Observations: 197
```

```
. genstest, test(phi) init(0 0.38) null(last) ci(0.50 0.99, points(20)) stab sb
Note: using the nuisS and/or varS options will decrease computation time for
> the single-break tests.
```

Test	Statistic	P-value	CI (alpha=.05)
S	2.712954	0.607	[.622, .99]
qLL-S	32.978783	0.310	[.647, .99]
ave-S	12.526473	0.218	[.72, .99]
exp-S	18.453388	0.077	[.818, .99]
sup-S	24.109671	0.082	[.818, .99]
qLL-stab-S	30.265829	0.234	[.573, .99]
ave-stab-S	9.813519	0.069	[.794, .99]
exp-stab-S	15.740433	0.025	[.867, .99]
sup-stab-S	21.396717	0.034	[.843, .99]

```
Tested null hypothesis vector: <phi> = < 0.821  >
Number of Instruments - Included Instruments: 6 - 2
Number of Observations: 197
```

The shrinkage in the gen-S confidence intervals relative to the S is due to the stability restrictions. This reduction is particularly remarkable for the exp-S test.

The next call of the `genstest` command illustrates how to test multiple parameters. In this call, we also set a grid search for estimating confidence regions for (ρ, ϕ) on the basis of the gen-S tests. The confidence regions are the collection of points $(\rho_0, \theta_0) \in (0,1) \times (0,1)$ in the grid search that do not reject the null hypothesis $H_0 : (\rho, \phi) = (\rho_0, \phi_0)$.

```
. genstest, test(rho phi) null(last) init(0.0) derivative(/g=-1)
> ci(0.01 .99 0.01 0.99, points(10 10)) stab sb
Note: using the nuisS and/or varS options will decrease computation time for
> the single-break tests.
```

Test	Statistic	P-value
S	3.043952	0.693
qLL-S	37.768561	0.136
ave-S	12.948586	0.271
exp-S	18.421490	0.110
sup-S	23.887305	0.117
qLL-stab-S	34.724609	0.068
ave-stab-S	9.904634	0.066
exp-stab-S	15.377538	0.029
sup-stab-S	20.843353	0.042

```
Tested null hypothesis vector: <rho phi> = < 0.386   0.821  >
Number of Instruments - Included Instruments: 6 - 1
Number of Observations: 197
```

Example 3 takes longer to run than examples 1 and 2 because the single-break tests are estimating recursively $\widehat{\gamma}(\theta, j)$, $\Phi_1(\theta)$, and $\Phi_2(\theta)$ 138 times for each point in the grid search.[7] Removing the `sb` option or adding `nuisS` and `varS` will reduce the computation time significantly.

The confidence sets for the joint hypothesis are shown below for the Wald and S tests. The graph for the Wald confidence set was created using built-in Stata commands (`test` and `gmm`). The other graph was created using the stored results of the `genstest` command. In the same graphs, we plot the confidence intervals of each parameter.

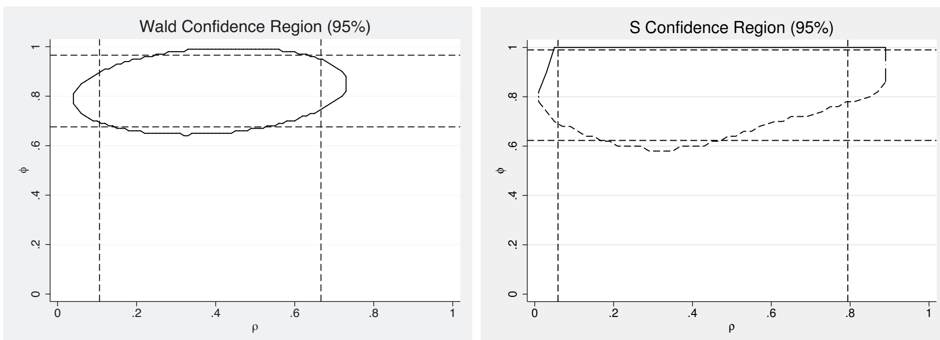

Figure 1. Wald and S 95%-level confidence sets for ϕ and ρ in the NKPC. The forcing variable is the log of the labor share. Instruments: constant, two lags of $\Delta\pi$, and three lags of x_t. Period: 1960q1–2008q3.

7. It takes approximately three to four minutes to compute confidence intervals for ρ and ϕ, respectively. In the case of the confidence regions, it takes 16 minutes to test all 100 points of the defined grid search. The reported times are obtained after executing the code in a PC with Intel(R) Core(TM) I7-2600 CPU 3.4 GHz processor with 4 GB RAM memory, Windows 7 system with Stata/IC 12.1.

Comparing the S confidence set with the results of the Wald test above, we might suspect the presence of weak instruments: the confidence interval of ρ generated by the S test covers almost the entire parameter space.

Next we obtain the confidence interval and regions using the ave-S, exp-S, sup-S, and qLL-S tests. The proposed tests, which are robust to weak instruments, generate smaller confidence intervals and regions than both the Wald and the S tests.

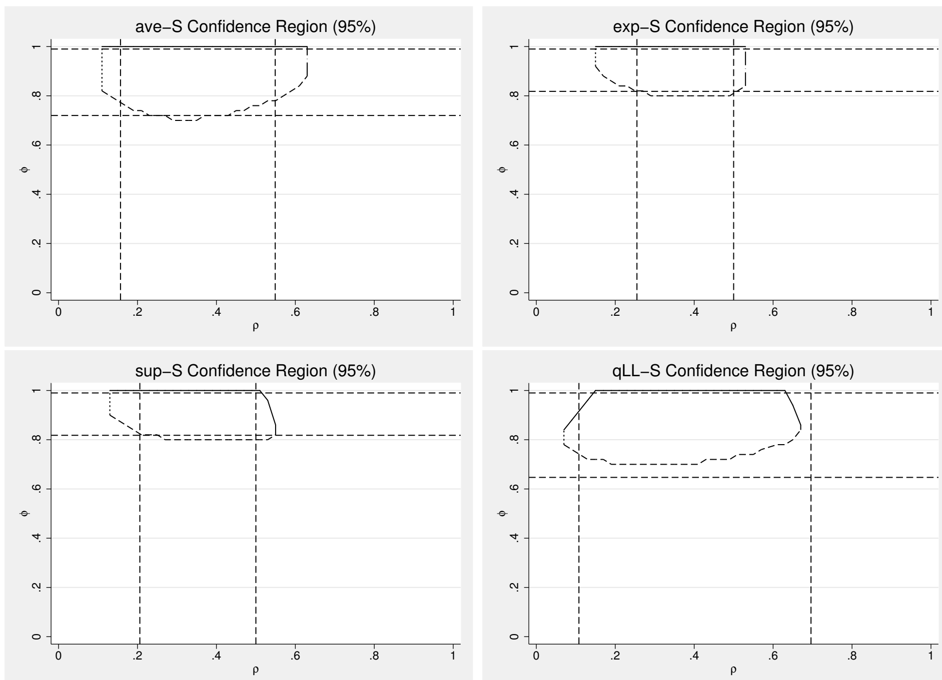

Figure 2. gen-S 95%-level confidence sets for ϕ and ρ in the NKPC. The forcing variable is the log of the labor share. Instruments: constant, two lags of $\Delta\pi$, and three lags of x_t. Period: 1960q1–2008q3.

We illustrate the importance of imposing the stability restrictions when performing inference in figure 3. The ave-S test is a combination of the S and ave-$\widetilde{S}$ tests [see (6)]. In the S-test confidence region graph, the confidence interval for ρ covers almost the entire parameter space. The range of points in the parameter space that satisfies the stability restrictions is a small fraction of the range of points that satisfies the S test, as illustrated by the ave-$\widetilde{S}$ confidence region graph. The same explanation can be extended to explaining the reduction in the range of ϕ.

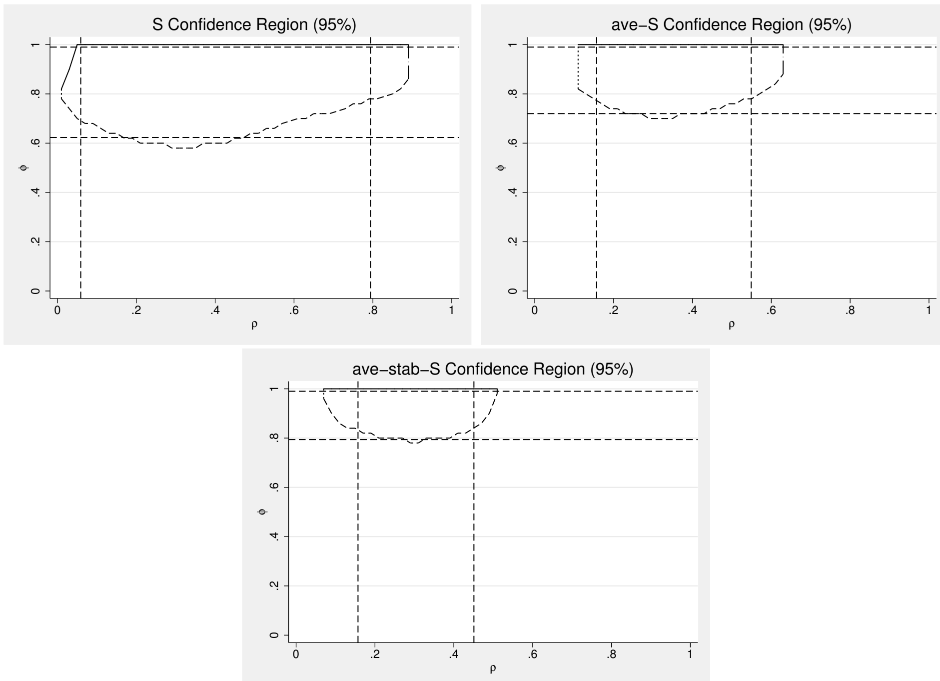

Figure 3. S, ave-S, and ave-$\widetilde{\mathrm{S}}$ 95%-level confidence sets for ϕ and ρ in the NKPC. The forcing variable is the log of the labor share. Instruments: constant, two lags of $\Delta\pi$, and three lags of x_t. Period: 1960q1–2008q3.

This last example illustrates that in time-series applications, the generalized S tests improve inference of structural parameters by incorporating information about instabilities in the moment condition without imposing identification restrictions implicitly assumed by computing the Wald test.

6 Acknowledgments

We would like to thank Mark Schaffer and Daniel Millimet for suggestions and comments. An anonymous referee helped us to improve the manuscript and the `genstest` code, and we are grateful for this. We also would like to thank the Tulane Research Experience for Undergraduates in Applied Microeconomics and Program Evaluation (Treu-Ampe) and the Provost's Fund for Faculty–Student Scholarly and Artistic Engagement for funding support.

7 References

Anderson, T. W., and H. Rubin. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20: 46–63.

Andrews, D. W. K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61: 821–856.

Angrist, J. D., and A. B. Krueger. 1995. Split-sample instrumental variables estimates of the return to schooling. *Journal of Business and Economic Statistics* 13: 225–235.

Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics Using Stata*. Rev. ed. College Station, TX: Stata Press.

Caner, M. 2007. Boundedly pivotal structural change tests in continuous updating GMM with strong, weak identification and completely unidentified cases. *Journal of Econometrics* 137: 28–67.

Chernozhukov, V., and C. Hansen. 2008. The reduced form: A simple approach to inference with weak instruments. *Economics Letters* 100: 68–71.

Davidson, R., and J. G. MacKinnon. 2003. *Econometric Theory and Methods*. New York: Oxford University Press.

Elliott, G., and U. K. Müller. 2006. Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies* 73: 907–940.

Hall, A. R. 2005. *Generalized Method of Moments*. Oxford: Oxford University Press.

Hall, A. R., and A. Sen. 1999. Structural stability testing in models estimated by generalized method of moments. *Journal of Business and Economic Statistics* 17: 335–348.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

Kleibergen, F., and S. Mavroeidis. 2009. Weak instrument robust tests in GMM and the new Keynesian Phillips curve. *Journal of Business and Economic Statistics* 27: 293–311.

Magnusson, L. M., and S. Mavroeidis. 2010a. Identification using stability restrictions. Working Paper 1116, Department of Economics, Tulane University. http://ideas.repec.org/p/tul/wpaper/1116.html.

———. 2010b. Identification-robust minimum distance estimation of the new Keynesian Phillips curve. *Journal of Money, Credit and Banking* 42: 465–481.

Mroz, T. A. 1987. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55: 765–799.

Newey, W. K., and K. D. West. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61: 631–653.

Piehl, A. M., S. J. Cooper, A. A. Braga, and D. M. Kennedy. 2003. Testing for structural breaks in the evaluation of programs. *Review of Economics and Statistics* 85: 550–558.

Sbordone, A. M. 2005. Do expected future marginal costs drive inflation dynamics? *Journal of Monetary Economics* 52: 1183–1197.

Sowell, F. 1996. Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* 64: 1085–1107.

StataCorp. 2009. *Stata: Release 11.* Statistical Software. College Station, TX: StataCorp LP.

Stock, J. H., and M. W. Watson. 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14: 11–30.

Stock, J. H., and J. H. Wright. 2000. GMM with weak identification. *Econometrica* 68: 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20: 518–529.

About the authors

Zachary Flynn is a PhD candidate in economics at the University of Wisconsin–Madison.

Leandro Magnusson is an assistant professor in the Department of Economics at the University of Western Australia.