

Powerful new tools for time series analysis

Christopher F Baum

Boston College & DIW

August 2007

This presentation discusses two recent developments in time series analysis by Graham Elliott of UCSD and coauthors and their implementations as Stata commands, available via `ssc`.

The first, `urcovar`, is a test for a unit root in a time series (Elliott & Michael Jansson, *Journal of Econometrics*, 2003) which extends the ERS `dfgls` test by adding stationary covariates to gain additional power.

The second, `qll`, is based on Elliott and Ulrich Müller's *Review of Economic Studies* paper (2006) in which they illustrate that tests for parameter constancy and tests for an unknown break process can be unified to produce a single efficient q_{LL} test for stability of the regression function.

This presentation discusses two recent developments in time series analysis by Graham Elliott of UCSD and coauthors and their implementations as Stata commands, available via `ssc`.

The first, `urcovar`, is a test for a unit root in a time series (Elliott & Michael Jansson, *Journal of Econometrics*, 2003) which extends the ERS `dfgls` test by adding stationary covariates to gain additional power.

The second, `qll`, is based on Elliott and Ulrich Müller's *Review of Economic Studies* paper (2006) in which they illustrate that tests for parameter constancy and tests for an unknown break process can be unified to produce a single efficient q_{LL} test for stability of the regression function.

This presentation discusses two recent developments in time series analysis by Graham Elliott of UCSD and coauthors and their implementations as Stata commands, available via `ssc`.

The first, `urcovar`, is a test for a unit root in a time series (Elliott & Michael Jansson, *Journal of Econometrics*, 2003) which extends the ERS `dfgls` test by adding stationary covariates to gain additional power.

The second, `qll`, is based on Elliott and Ulrich Müller's *Review of Economic Studies* paper (2006) in which they illustrate that tests for parameter constancy and tests for an unknown break process can be unified to produce a single efficient q_{LL} test for stability of the regression function.

urcovar

The presence of a unit root in the time series representation of a variable has important implications for both the econometric method used and the economic interpretation of the model in which that variable appears.

The “first generation” tests of Dickey and Fuller (`dfuller`) and Phillips and Perron (`pperron`) have been supplanted by more powerful “second generation” tests such as `dfgls` of Elliott–Rothenberg–Stock. ERS showed that there is no uniformly most powerful test, but that their DF-GLS alternative is approximately most powerful for the unit root testing problem.

urcovar

The presence of a unit root in the time series representation of a variable has important implications for both the econometric method used and the economic interpretation of the model in which that variable appears.

The “first generation” tests of Dickey and Fuller (`dfuller`) and Phillips and Perron (`pperron`) have been supplanted by more powerful “second generation” tests such as `dfgls` of Elliott–Rothenberg–Stock. ERS showed that there is no uniformly most powerful test, but that their DF-GLS alternative is approximately most powerful for the unit root testing problem.

As is well known in the applied economics literature, even a test with DF-GLS's favorable characteristics may still lack power to distinguish between the null hypothesis of nonstationary behavior ($I(1)$) and the stationary alternative ($I(0)$).

In many applications, using a longer time series is not feasible due to known structural breaks, institutional changes, and the like. Another potential alternative, the panel unit root test, brings its own set of complications (do we assume all series are $I(1)$? that all are $I(0)$?).

As is well known in the applied economics literature, even a test with DF-GLS's favorable characteristics may still lack power to distinguish between the null hypothesis of nonstationary behavior ($I(1)$) and the stationary alternative ($I(0)$).

In many applications, using a longer time series is not feasible due to known structural breaks, institutional changes, and the like. Another potential alternative, the panel unit root test, brings its own set of complications (do we assume all series are $I(1)$? that all are $I(0)$?).

Elliott and Jansson addressed this issue by considering a model in which there is one potentially nonstationary ($I(1)$) series, y , which potentially covaries with some available stationary variables, x . This idea was first put forth by Bruce Hansen in 1995 who proposed a covariate augmented D–F test, or CADF test and showed that this test had greater power than those which ignored the covariates.

Elliott and Jansson extended Hansen's results to show that such a test could be conducted in the presence of unknown nuisance parameters, and with constants and trends in the model. Their proposed test may be readily calculated by estimating a vector autoregression (var) in the $\{y, x\}$ variables and performing a sequence of matrix manipulations.

Elliott and Jansson addressed this issue by considering a model in which there is one potentially nonstationary ($I(1)$) series, y , which potentially covaries with some available stationary variables, x . This idea was first put forth by Bruce Hansen in 1995 who proposed a covariate augmented D–F test, or CADF test and showed that this test had greater power than those which ignored the covariates.

Elliott and Jansson extended Hansen's results to show that such a test could be conducted in the presence of unknown nuisance parameters, and with constants and trends in the model. Their proposed test may be readily calculated by estimating a vector autoregression (var) in the $\{y, x\}$ variables and performing a sequence of matrix manipulations.

The model considered is:

$$z_t = \beta_0 + \beta_1 t + u_t, \quad t = 1, \dots, T$$

$$A(L) \begin{pmatrix} (1 - \rho L)u_{y,t} \\ u_{x,t} \end{pmatrix} = e_t$$

with $z_t = \{y_t, x_t'\}'$, x_t an $m \times 1$ vector,

$\beta_0 = \{\beta_{y0}, \beta'_{x0}\}'$, $\beta_1 = \{\beta_{y1}, \beta'_{x1}\}'$ and $u_t = \{u_{y,t}, u'_{x,t}\}'$.

$A(L)$ is a stable matrix polynomial of finite order k in the lag operator L .

This is a vector autoregression (VAR) in the model of x and the quasi-difference of y . The relevant test is that the parameter ρ is equal to unity, implying that y has a unit root, against alternatives that ρ is less than one.

The model considered is:

$$z_t = \beta_0 + \beta_1 t + u_t, \quad t = 1, \dots, T$$

$$A(L) \begin{pmatrix} (1 - \rho L)u_{y,t} \\ u_{x,t} \end{pmatrix} = e_t$$

with $z_t = \{y_t, x_t'\}'$, x_t an $m \times 1$ vector,

$\beta_0 = \{\beta_{y0}, \beta'_{x0}\}'$, $\beta_1 = \{\beta_{y1}, \beta'_{x1}\}'$ and $u_t = \{u_{y,t}, u'_{x,t}\}'$.

$A(L)$ is a stable matrix polynomial of finite order k in the lag operator L .

This is a vector autoregression (VAR) in the model of x and the quasi-difference of y . The relevant test is that the parameter ρ is equal to unity, implying that y has a unit root, against alternatives that ρ is less than one.

The potential gain in this test depends on the R^2 between y and the set of x covariates. As Elliott and Pesavento (*J. Money, Credit, Banking* 2006) point out, the relevant issue is the ability of a unit root test to have power to distinguish between $I(1)$ and a local alternative. The local alternative is in terms of $c = T(\rho - 1)$ where ρ is the largest root in the AR representation of y .

How far below unity must ρ fall to give a unit root test the ability to discern stationary, mean-reverting behavior (albeit with strong persistence, with $\rho > 0.9$) from nonstationary, unit root behavior? The various tests in this literature differ in their power against relevant local alternatives.

The potential gain in this test depends on the R^2 between y and the set of x covariates. As Elliott and Pesavento (*J. Money, Credit, Banking* 2006) point out, the relevant issue is the ability of a unit root test to have power to distinguish between $I(1)$ and a local alternative. The local alternative is in terms of $c = T(\rho - 1)$ where ρ is the largest root in the AR representation of y .

How far below unity must ρ fall to give a unit root test the ability to discern stationary, mean-reverting behavior (albeit with strong persistence, with $\rho > 0.9$) from nonstationary, unit root behavior? The various tests in this literature differ in their power against relevant local alternatives.

The c parameter can be expressed in terms of the half-life (k) of a shock, where a unit root implies an infinite half-life:

$k = \log(0.5) / \log(\rho)$. When related to the local alternative, $k/T = \log(0.5)/c$. For about 120 observations (30 years of quarterly data), $c = -5$ corresponds to a half-life of 16.8 time periods (over four years at a quarterly frequency). From Elliott, Rothenberg, Stock (ERS, *Econometrica*, 1996), the standard Dickey-Fuller test (`dfuller`) has 12% power to reject the alternative. The ERS DF-GLS test (`dfgls`) has 32% power.

In contrast, with an $R^2 = 0.2$, the Elliott–Jansson (EJ) test has power of 42%. The power rises to 53% (69%) for $R^2 = 0.4$ (0.6). For higher absolute values of c (shorter half-lives), the gains are smaller. For $c = -10$, or a half-life of 8.4 periods, the power of the D-F (DF-GLS) test is 31% (75%). The EJ test has power of 88%, 94% and 99% for $R^2 = 0.2, 0.4, 0.6$. One clear conclusion: DF-GLS always has superior power compared to `dfuller`.

The c parameter can be expressed in terms of the half-life (k) of a shock, where a unit root implies an infinite half-life:

$k = \log(0.5) / \log(\rho)$. When related to the local alternative, $k/T = \log(0.5) / c$. For about 120 observations (30 years of quarterly data), $c = -5$ corresponds to a half-life of 16.8 time periods (over four years at a quarterly frequency). From Elliott, Rothenberg, Stock (ERS, *Econometrica*, 1996), the standard Dickey-Fuller test (`dfuller`) has 12% power to reject the alternative. The ERS DF-GLS test (`dfgls`) has 32% power.

In contrast, with an $R^2 = 0.2$, the Elliott–Jansson (EJ) test has power of 42%. The power rises to 53% (69%) for $R^2 = 0.4$ (0.6). For higher absolute values of c (shorter half-lives), the gains are smaller. For $c = -10$, or a half-life of 8.4 periods, the power of the D-F (DF-GLS) test is 31% (75%). The EJ test has power of 88%, 94% and 99% for $R^2 = 0.2, 0.4, 0.6$. One clear conclusion: DF-GLS always has superior power compared to `dfuller`.

Like other unit root tests, you must specify the deterministic model assumed for y . As in `dfgls`, you can specify that the model contains no deterministic terms, a constant, or constant and trend. But as the Elliott–Jansson model contains one or more x variables as well, you may also specify that those variables' deterministic model contains no deterministic terms, a constant, or constant and trend.

Five cases are defined:

- 1 No constant nor trend in model
- 2 Constant in y only
- 3 Constants in both $\{y, x\}$
- 4 Constant and trend in y , constant in x
- 5 No restrictions

Like other unit root tests, you must specify the deterministic model assumed for y . As in `dfgls`, you can specify that the model contains no deterministic terms, a constant, or constant and trend. But as the Elliott–Jansson model contains one or more x variables as well, you may also specify that those variables' deterministic model contains no deterministic terms, a constant, or constant and trend.

Five cases are defined:

- 1 No constant nor trend in model
- 2 Constant in y only
- 3 Constants in both $\{y, x\}$
- 4 Constant and trend in y , constant in x
- 5 No restrictions

Like other unit root tests, you must specify the deterministic model assumed for y . As in `dfgls`, you can specify that the model contains no deterministic terms, a constant, or constant and trend. But as the Elliott–Jansson model contains one or more x variables as well, you may also specify that those variables' deterministic model contains no deterministic terms, a constant, or constant and trend.

Five cases are defined:

- 1 No constant nor trend in model
- 2 Constant in y only
- 3 Constants in both $\{y, x\}$
- 4 Constant and trend in y , constant in x
- 5 No restrictions

Like other unit root tests, you must specify the deterministic model assumed for y . As in `dfgls`, you can specify that the model contains no deterministic terms, a constant, or constant and trend. But as the Elliott–Jansson model contains one or more x variables as well, you may also specify that those variables' deterministic model contains no deterministic terms, a constant, or constant and trend.

Five cases are defined:

- 1 No constant nor trend in model
- 2 Constant in y only
- 3 Constants in both $\{y, x\}$
- 4 Constant and trend in y , constant in x
- 5 No restrictions

Like other unit root tests, you must specify the deterministic model assumed for y . As in `dfgls`, you can specify that the model contains no deterministic terms, a constant, or constant and trend. But as the Elliott–Jansson model contains one or more x variables as well, you may also specify that those variables' deterministic model contains no deterministic terms, a constant, or constant and trend.

Five cases are defined:

- 1 No constant nor trend in model
- 2 Constant in y only
- 3 Constants in both $\{y, x\}$
- 4 Constant and trend in y , constant in x
- 5 No restrictions

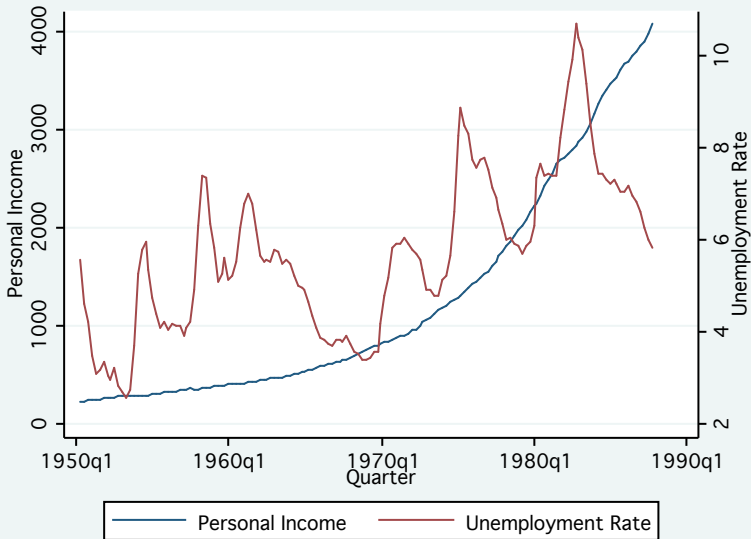
The `urcovar` Stata command implements the EJ test using Mata to perform a complicated sequence of matrix manipulations that produce the test statistic. EJ's Table 1 of asymptotic critical values is stored in the program and used to produce a critical value corresponding to the R^2 for your data. The command syntax:

```
urcovar depvar varlist [if exp] [in range] [ , maxlag(#) case(#) firstobs ]
```

where the `case` option specifies the deterministic model, with default of case 1. The `maxlag` option specifies the number of lags to be used in computing the VAR (default 1). The `firstobs` option specifies that the first observation of `depvar` should be used to define the first quasi-difference (rather than zero). The `urcovar` command is available for Stata 9.2 or Stata 10 from `ssc`.

As an illustration of `urcovar` use, we consider an experiment similar to that tested in EJ, who in turn refer to a Blanchard–Quah model. The variable of interest is U.S. personal income (`PINCOME`). The single stationary covariate to be considered is the U.S. unemployment rate (`UNRATE`). Both are acquired from the FRED database with the `freduse` command (Drukker, *Stata Journal*, 2006) and have been converted to the common quarterly frequency for 1950Q2–1987Q4 using `tscollap` (Baum, *STB-57*, 2000).

We first present a line plot of these two series, then the output from a conventional DF-GLS test, followed by the output from `urcovar`, cases 3 and 5. The `maxlag` considered in both the `dfgls` and `urcovar` tests is set to eight quarters.




```
. dfglr PINCOME, maxlag(8) trend
```

```
DF-GLS for PINCOME
```

```
Number of obs = 142
```

[lags]	DF-GLS tau Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
8	-0.830	-3.519	-2.875	-2.593
7	-0.585	-3.519	-2.890	-2.607
6	-0.173	-3.519	-2.905	-2.620
5	-0.071	-3.519	-2.918	-2.633
4	-0.272	-3.519	-2.932	-2.645
3	-0.334	-3.519	-2.944	-2.656
2	0.332	-3.519	-2.955	-2.666
1	0.655	-3.519	-2.966	-2.676

```
Opt Lag (Ng-Perron seq t) = 8 with RMSE 11.44531
```

```
Min SC = 5.103851 at lag 3 with RMSE 11.96667
```

```
Min MAIC = 5.000489 at lag 7 with RMSE 11.56349
```

```
. urcovar PINCOME UNRATE, maxlag(8) case(3)
Elliott-Jansson unit root test for PINCOME      1950q2 - 1987q4
Number of obs:      143
Stationary covariates: UNRATE
Deterministic model: Case 3
Maximum lag order:   8
Estimated R-squared: 0.9950
H0: rho = 1 [ PINCOME is I(1) ]
H1: rho < 1 [ PINCOME is I(0) ]
Reject H0 if Lambda < critical value
Lambda:              11.8307
5% critical value:   17.9900
```

```
. urcovar PINCOME UNRATE, maxlag(8) case(5)
Elliott-Jansson unit root test for PINCOME      1950q2 - 1987q4
Number of obs:      143
Stationary covariates: UNRATE
Deterministic model: Case 5
Maximum lag order:   8
Estimated R-squared: 0.8407
H0: rho = 1 [ PINCOME is I(1) ]
H1: rho < 1 [ PINCOME is I(0) ]
Reject H0 if Lambda < critical value
Lambda:              17.1558
5% critical value:   29.1170
```

The DF-GLS test is unable to reject its null of $I(1)$ at any reasonable level of significance. When we augment the test with the stationary covariate in the EJ test, quite different results are forthcoming. Case 3 allows for constant terms (but no trends) in both the quasi-difference of $PINCOME$ and $UNRATE$. The R^2 in this system is over 0.99. Case 5 allows constant terms and trends in both equations of the VAR, with an R^2 of 0.84.

Like the DF-GLS test, the EJ test has a null hypothesis of nonstationarity ($I(1)$). The EJ test statistic, λ , must be compared with the interpolated 5% critical value. A value of λ **smaller** than the tabulated value leads to a rejection, and *vice versa*. In both cases, we may reject the null hypothesis at the 95% level of confidence in favor of the alternative hypothesis of stationarity.

The DF-GLS test is unable to reject its null of $I(1)$ at any reasonable level of significance. When we augment the test with the stationary covariate in the EJ test, quite different results are forthcoming. Case 3 allows for constant terms (but no trends) in both the quasi-difference of $PINCOME$ and $UNRATE$. The R^2 in this system is over 0.99. Case 5 allows constant terms and trends in both equations of the VAR, with an R^2 of 0.84.

Like the DF-GLS test, the EJ test has a null hypothesis of nonstationarity ($I(1)$). The EJ test statistic, λ , must be compared with the interpolated 5% critical value. A value of λ **smaller** than the tabulated value leads to a rejection, and *vice versa*. In both cases, we may reject the null hypothesis at the 95% level of confidence in favor of the alternative hypothesis of stationarity.

Elliott and Müller's 2006 paper in *Review of Economic Studies* (EM) addresses the large literature on testing a time series model for structural stability. They consider “tests of the null hypothesis of a stable linear model

$$y_t = X_t' \bar{\beta} + Z_t' \gamma + \epsilon$$

against the alternative of a partially unstable model

$$y_t = X_t' \beta_t + Z_t' \gamma + \epsilon$$

where the variation in β_t is of the strong form” (p. 907), or nontrivial.

Consideration of this alternative has led to a huge literature based on the “diversity of possible ways $\{\beta_t\}$ can be non-constant.” EM point out that optimal tests and their asymptotic distributions have not been derived for many particular models of the alternative.

Their approach develops a single unified framework, noting that the “seemingly different approaches of ‘structural breaks’ and ‘random coefficients’ are in fact equivalent.” (p.908) EM unify the approaches that describe a breaking process with a number of non-random parameters with tests that specify stochastic processes for $\{\beta_t\}$ without requiring to specify its exact evolution.

The processes considered include breaks that occur in a random fashion, serial correlation in the changes of the coefficients, a clustering of break dates, and so on. Under a normality assumption on the disturbances, “small sample efficient tests in this broad set are asymptotically equivalent” and “leaving the exact breaking process unspecified (apart from a scaling parameter) does not result in a loss of power in large samples.” (p. 908)

Consideration of this alternative has led to a huge literature based on the “diversity of possible ways $\{\beta_t\}$ can be non-constant.” EM point out that optimal tests and their asymptotic distributions have not been derived for many particular models of the alternative.

Their approach develops a single unified framework, noting that the “seemingly different approaches of ‘structural breaks’ and ‘random coefficients’ are in fact equivalent.” (p.908) EM unify the approaches that describe a breaking process with a number of non-random parameters with tests that specify stochastic processes for $\{\beta_t\}$ without requiring to specify its exact evolution.

The processes considered include breaks that occur in a random fashion, serial correlation in the changes of the coefficients, a clustering of break dates, and so on. Under a normality assumption on the disturbances, “small sample efficient tests in this broad set are asymptotically equivalent” and “leaving the exact breaking process unspecified (apart from a scaling parameter) does not result in a loss of power in large samples.” (p. 908)

Consideration of this alternative has led to a huge literature based on the “diversity of possible ways $\{\beta_t\}$ can be non-constant.” EM point out that optimal tests and their asymptotic distributions have not been derived for many particular models of the alternative.

Their approach develops a single unified framework, noting that the “seemingly different approaches of ‘structural breaks’ and ‘random coefficients’ are in fact equivalent.” (p.908) EM unify the approaches that describe a breaking process with a number of non-random parameters with tests that specify stochastic processes for $\{\beta_t\}$ without requiring to specify its exact evolution.

The processes considered include breaks that occur in a random fashion, serial correlation in the changes of the coefficients, a clustering of break dates, and so on. Under a normality assumption on the disturbances, “small sample efficient tests in this broad set are asymptotically equivalent” and “leaving the exact breaking process unspecified (apart from a scaling parameter) does not result in a loss of power in large samples.” (p. 908)

The consequences of this approach to the problem of structural stability are profound. “The equivalence of power over many models means that there is little point in deriving further optimal tests for particular processes in our set” (p. 908) and the researcher can carry out (almost) efficient inference without specifying the exact path of the breaking process.

Furthermore, the computation of EM’s Quasi-Local Level (q_{LL}) test statistic is straightforward, and it remains valid for very general specifications of the error term and covariates. The computation requires no more than $(k + 1)$ OLS regressions for a model with k covariates, in contrast to many approaches which require T or T^2 regressions. No arbitrary trimming of the data is required.

The consequences of this approach to the problem of structural stability are profound. “The equivalence of power over many models means that there is little point in deriving further optimal tests for particular processes in our set” (p. 908) and the researcher can carry out (almost) efficient inference without specifying the exact path of the breaking process.

Furthermore, the computation of EM’s Quasi-Local Level (q_{LL}) test statistic is straightforward, and it remains valid for very general specifications of the error term and covariates. The computation requires no more than $(k + 1)$ OLS regressions for a model with k covariates, in contrast to many approaches which require T or T^2 regressions. No arbitrary trimming of the data is required.

In the *structural break* literature, a fixed number of N breaks at τ_1, \dots, τ_N are assumed. Much of the literature addresses $N = 1$: e.g. the “Chow test”, cusums tests of Brown–Durbin–Evans, Bai and Perron, Andrews and Ploberger, etc.

In contrast, the *time-varying parameter* literature considers a random process generating β_t : often considered as a random walk process. The approaches of Leybourne and McCabe, Nyblom, and Saikkonen and Luukonen are based on classical statistics, while Koop and Potter and Giordani et al. consider a Bayesian approach. All of these approaches are very analytically challenging.

EM argue that tests for one of these phenomena will have power against the other, and *vice versa*. Therefore a single approach will suffice.

In the *structural break* literature, a fixed number of N breaks at τ_1, \dots, τ_N are assumed. Much of the literature addresses $N = 1$: e.g. the “Chow test”, cusums tests of Brown–Durbin–Evans, Bai and Perron, Andrews and Ploberger, etc.

In contrast, the *time-varying parameter* literature considers a random process generating β_t : often considered as a random walk process. The approaches of Leybourne and McCabe, Nyblom, and Saikkonen and Luukonen are based on classical statistics, while Koop and Potter and Giordani et al. consider a Bayesian approach. All of these approaches are very analytically challenging.

EM argue that tests for one of these phenomena will have power against the other, and *vice versa*. Therefore a single approach will suffice.

In the *structural break* literature, a fixed number of N breaks at τ_1, \dots, τ_N are assumed. Much of the literature addresses $N = 1$: e.g. the “Chow test”, cusums tests of Brown–Durbin–Evans, Bai and Perron, Andrews and Ploberger, etc.

In contrast, the *time-varying parameter* literature considers a random process generating β_t : often considered as a random walk process. The approaches of Leybourne and McCabe, Nyblom, and Saikkonen and Luukonen are based on classical statistics, while Koop and Potter and Giordani et al. consider a Bayesian approach. All of these approaches are very analytically challenging.

EM argue that tests for one of these phenomena will have power against the other, and *vice versa*. Therefore a single approach will suffice.

EM raise the interesting question: why do we test for parameter constancy? They consider three motivations:

- 1 Stability relates to theoretical constructs such as the Lucas critique of economic policymaking
- 2 Forecasting will depend crucially on a stable relationship
- 3 Standard inference on $\bar{\beta}$ will be useless if $\{\beta_t\}$ varies in a permanent fashion; persistent changes will render a fixed model misleading

“The more pervasive these three motivations are, the more persistent the changes in $\{\beta\}$.” (p. 912) Therefore, EM propose that a useful test should maximize its power against persistent changes in $\{\beta_t\}$.

EM raise the interesting question: why do we test for parameter constancy? They consider three motivations:

- 1 Stability relates to theoretical constructs such as the Lucas critique of economic policymaking
- 2 Forecasting will depend crucially on a stable relationship
- 3 Standard inference on $\bar{\beta}$ will be useless if $\{\beta_t\}$ varies in a permanent fashion; persistent changes will render a fixed model misleading

“The more pervasive these three motivations are, the more persistent the changes in $\{\beta\}$.” (p. 912) Therefore, EM propose that a useful test should maximize its power against persistent changes in $\{\beta_t\}$.

EM raise the interesting question: why do we test for parameter constancy? They consider three motivations:

- 1 Stability relates to theoretical constructs such as the Lucas critique of economic policymaking
- 2 Forecasting will depend crucially on a stable relationship
- 3 Standard inference on $\bar{\beta}$ will be useless if $\{\beta_t\}$ varies in a permanent fashion; persistent changes will render a fixed model misleading

“The more pervasive these three motivations are, the more persistent the changes in $\{\beta\}$.” (p. 912) Therefore, EM propose that a useful test should maximize its power against persistent changes in $\{\beta_t\}$.

EM raise the interesting question: why do we test for parameter constancy? They consider three motivations:

- 1 Stability relates to theoretical constructs such as the Lucas critique of economic policymaking
- 2 Forecasting will depend crucially on a stable relationship
- 3 Standard inference on $\bar{\beta}$ will be useless if $\{\beta_t\}$ varies in a permanent fashion; persistent changes will render a fixed model misleading

“The more pervasive these three motivations are, the more persistent the changes in $\{\beta\}$.” (p. 912) Therefore, EM propose that a useful test should maximize its power against persistent changes in $\{\beta_t\}$.

The conditions underlying the EM test allow for diverse breaking models, from relatively rare (including a single break) to very frequent small breaks (such as breaks every period with probability p). Breaks can also occur with a regular pattern, such as every 16 quarters following U.S. presidential elections.

Computation of the q_{LL} test statistic is straightforward, relying only on OLS regressions and construction of an estimate of the long-run covariance matrix of $\{X_t\epsilon_t\}$. For uncorrelated ϵ_t , a robust covariance matrix will suffice. For possibly autocorrelated ϵ_t , a HAC (Newey–West) covariance matrix is appropriate.

The null hypothesis of parameter stability is rejected for **small** values of \widehat{q}_{LL} : that is, values more negative than the critical values. Asymptotic critical values are provided by EM for $k = 1, \dots, 10$ and are independent of the dimension of Z_t (the set of covariates assumed to have stable coefficients).

The conditions underlying the EM test allow for diverse breaking models, from relatively rare (including a single break) to very frequent small breaks (such as breaks every period with probability p). Breaks can also occur with a regular pattern, such as every 16 quarters following U.S. presidential elections.

Computation of the q_{LL} test statistic is straightforward, relying only on OLS regressions and construction of an estimate of the long-run covariance matrix of $\{X_t\epsilon_t\}$. For uncorrelated ϵ_t , a robust covariance matrix will suffice. For possibly autocorrelated ϵ_t , a HAC (Newey–West) covariance matrix is appropriate.

The null hypothesis of parameter stability is rejected for **small** values of \widehat{q}_{LL} : that is, values more negative than the critical values. Asymptotic critical values are provided by EM for $k = 1, \dots, 10$ and are independent of the dimension of Z_t (the set of covariates assumed to have stable coefficients).

The conditions underlying the EM test allow for diverse breaking models, from relatively rare (including a single break) to very frequent small breaks (such as breaks every period with probability p). Breaks can also occur with a regular pattern, such as every 16 quarters following U.S. presidential elections.

Computation of the q_{LL} test statistic is straightforward, relying only on OLS regressions and construction of an estimate of the long-run covariance matrix of $\{X_t\epsilon_t\}$. For uncorrelated ϵ_t , a robust covariance matrix will suffice. For possibly autocorrelated ϵ_t , a HAC (Newey–West) covariance matrix is appropriate.

The null hypothesis of parameter stability is rejected for **small** values of \widehat{q}_{LL} : that is, values more negative than the critical values. Asymptotic critical values are provided by EM for $k = 1, \dots, 10$ and are independent of the dimension of Z_t (the set of covariates assumed to have stable coefficients).

The `qll` Stata command implements the EM q_{LL} test using Mata to produce the test statistic. EM's Table 1 of asymptotic critical values is stored in the program and used to produce 10%, 5% and 1% critical values corresponding to number of regressors with potentially unstable parameters. The command syntax:

```
qll depvar varlist [if exp] [in range] [ , (zvarlist) rlag(#) ]
```

where the parenthesized `zvarlist` optionally specifies the list of covariates assumed to have stable coefficients (none are required). The `rlag` option specifies the number of lags to be used in computing the long-run covariance matrix of $\{X_{t\epsilon_t}\}$. If a negative value is given, the optimal lag order is chosen by the BIC criterion. The `qll` command is available for Stata 9.2 or Stata 10 from `ssc`.

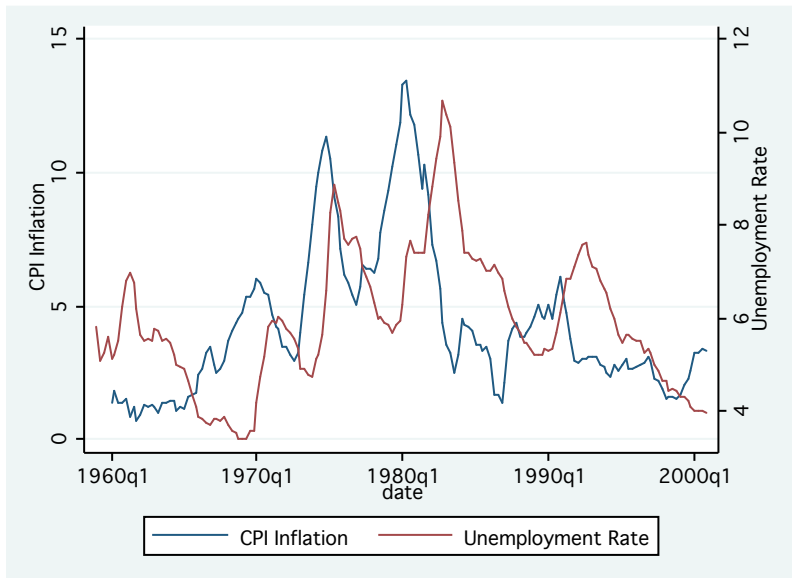
We consider a regression of inflation on the lagged unemployment rate, the Treasury bill rate and the Treasury bond rate. We assume the latter two coefficients are stable over the period. We test over the full sample and a 1990–2000 subsample.

The `qll` Stata command implements the EM q_{LL} test using Mata to produce the test statistic. EM's Table 1 of asymptotic critical values is stored in the program and used to produce 10%, 5% and 1% critical values corresponding to number of regressors with potentially unstable parameters. The command syntax:

```
qll depvar varlist [if exp] [in range] [ , (zvarlist) rlag(#) ]
```

where the parenthesized `zvarlist` optionally specifies the list of covariates assumed to have stable coefficients (none are required). The `rlag` option specifies the number of lags to be used in computing the long-run covariance matrix of $\{X_{t\epsilon_t}\}$. If a negative value is given, the optimal lag order is chosen by the BIC criterion. The `qll` command is available for Stata 9.2 or Stata 10 from `ssc`.

We consider a regression of inflation on the lagged unemployment rate, the Treasury bill rate and the Treasury bond rate. We assume the latter two coefficients are stable over the period. We test over the full sample and a 1990–2000 subsample.



```
. qll inf L.UR (TBILL TBON), rlag(8)
Elliott-Müller qLL test statistic for time varying coefficients
in the model of inf, 1960q1 - 2000q4
Allowing for time variation in 1 regressors
H0: all regression coefficients fixed over the sample period (N = 164)
Test stat.    1% Crit.Val.    5% Crit.Val.    10% Crit.Val.
      -2.260      -11.05      -8.36      -7.14
Long-run variance computed with 8 lags.

. qll inf L.UR (TBILL TBON) if tin(1990q1,), rlag(8)
Elliott-Müller qLL test statistic for time varying coefficients
in the model of inf, 1990q1 - 2000q4
Allowing for time variation in 1 regressors
H0: all regression coefficients fixed over the sample period (N = 44)
Test stat.    1% Crit.Val.    5% Crit.Val.    10% Crit.Val.
      -6.647      -11.05      -8.36      -7.14
Long-run variance computed with 8 lags.
```


In both samples, using eight lags to calculate the long-run covariance matrix, the null hypothesis that the coefficients on the lagged unemployment rate ($L.UR$) are stable cannot be rejected at the 10% level of confidence. The Elliott–Müller q_{LL} test indicates that the stability of this regression model, allowing for instability in the coefficient of the unemployment rate only, cannot be rejected by the data.

Acknowledgements

I am grateful to Graham Elliott for his cogent presentation of his techniques during his visit to Boston College in spring 2007, for helpful discussions of the methodology and for his MATLAB code implementing the `q11` routine. Remaining errors are my own.