CREATES Research Paper 2010-59

# Numerical distribution functions of fractional unit root and cointegration tests

James G. MacKinnon and Morten Ørregaard Nielsen

# Numerical distribution functions of fractional unit root and cointegration tests[*]

James G. MacKinnon[†]
Queen's University

Morten Ørregaard Nielsen
Queen's University and CREATES

September 10, 2010

## Abstract

We calculate numerically the asymptotic distribution functions of likelihood ratio tests for fractional unit roots and cointegration rank. Because these distributions depend on a real-valued parameter, $b$, which must be estimated, simple tabulation is not feasible. Partly due to the presence of this parameter, the choice of model specification for the response surface regressions used to obtain the numerical distribution functions is more involved than is usually the case. We deal with model uncertainty by model averaging rather than by model selection. We make available a computer program which, given the dimension of the problem, $q$, and a value of $b$, provides either a set of critical values or the asymptotic $P$ value for any value of the likelihood ratio statistic. The use of this program is illustrated by means of an empirical example involving opinion poll data.

JEL Codes: C12, C16, C22, C32

Keywords: cofractional process, fractional unit root, fractional cointegration, response surface regression, cointegration rank, numerical distribution function, model averaging

## 1  Introduction

This paper deals with calculation of numerical distribution functions — and hence critical values and $P$ values — of tests for fractional unit roots and cointegration based on the fractional vector autoregressive (FVAR) model proposed in Johansen (2008) and analyzed by

---

[†]Corresponding author. Address: Department of Economics, Queen's University, Dunning Hall, 94 University Avenue, Kingston, ON K7L 3N6, Canada. Email: jgm@econ.queensu.ca

Johansen and Nielsen (2010a, 2010b), henceforth JN. This model is given in error correction form as

$$\Delta^d \boldsymbol{X}_t = \Delta^{d-b} \boldsymbol{\Pi} L_b \boldsymbol{X}_t + \sum_{i=1}^{k} \boldsymbol{\Gamma}_i \Delta^d L_b^i \boldsymbol{X}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \tag{1}$$

where $\boldsymbol{\varepsilon}_t$ is $i.i.d.(\boldsymbol{0}, \boldsymbol{\Omega})$, $d \geq b > 0.5$, $\Delta^d$ is the fractional difference operator, and $L_b = 1 - \Delta^b$ is the fractional lag operator. The time series $\boldsymbol{X}_t$ is of dimension $p$. The model (1) includes the Johansen (1988) cointegrated VAR model as the special case $d = b = 1$, and the interpretation of the model parameters is similar. Thus, we can write $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where the $p \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with $r \leq p$, are assumed to have full column rank $r$. Then the columns of $\boldsymbol{\beta}$ are the $r$ cointegrating (cofractional) relations that determine the long-run equilibria, and $\boldsymbol{\alpha}$ holds the adjustment coefficients. The rank $r$ is termed the cointegrating, or cofractional, rank. Note that the expansion of $L_b = 1 - \Delta^b$ has no term in $L^0$, and thus only lagged disequilibrium errors appear in (1).

Readers will note that the model (1) does not contain any deterministic terms, because the results of JN(2010a, 2010b) do not cover FVAR models with deterministic terms. This is an area of ongoing research.

JN (2010b) shows how to estimate the parameters of the model (1), including the fractional orders $b$ and $d$, and how to test hypotheses on the cointegration rank $r$. The tests include, as special cases, Johansen's (1988) tests for cointegration in VARs (when $d = b = 1$), standard Dickey-Fuller tests (when $d = b = 1$ and $q = 1$), and fractional Dickey-Fuller tests (JN, 2010a) (when $q = 1$).

Programs for computing accurate asymptotic critical values and asymptotic $P$ values, based on numerical distribution functions (NDFs), are available for standard Dickey-Fuller tests (MacKinnon, 1996) and for Johansen tests (MacKinnon, Haug, and Michelis, 1999). Our objective is to obtain comparable NDFs for tests of cointegration rank in on the fractional model (1).

In fractionally integrated models, a new complication is the dependence of the asymptotic distributions on a real-valued parameter, $b$, which implies that critical values have to be simulated conditional on an estimate of that parameter. An additional complication that arises partially because of the presence of the parameter $b$ is that the choice of model specification for the response surface regressions is more involved than is the case for ordinary unit root and cointegration tests. One methodological innovation of this paper is that we use model averaging rather than model selection to deal with model uncertainty in the specification of response surface regressions.

The remainder of the paper is organized as follows. Section 2 briefly introduces the tests and associated asymptotic distributions. Section 3 discusses how we obtain numerical approximations to the distribution functions of the tests. Section 4 explains how our results can be used to obtain approximate asymptotic critical values and $P$ values via computer programs that we make available. Section 5 presents an empirical example that uses polling data for the United Kingdom, and Section 6 concludes.

## 2  Likelihood ratio tests for fractional cointegration

JN (2010b) shows that, under i.i.d errors with suitable moment conditions, the maximum likelihood parameter estimates $(\hat{d}, \hat{b}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Gamma}}_1, \ldots, \hat{\boldsymbol{\Gamma}}_k)$ are asymptotically Gaussian, while the estimated cointegration vectors in $\hat{\boldsymbol{\beta}}$ are locally asymptotically mixed normal. These results allow asymptotically standard (chi-squared) inference on all parameters of the model — including the cointegrating relations and orders of fractionality — using Gaussian likelihood ratio tests.

In this paper, we are concerned with the asymptotic distribution of the likelihood ratio (LR) test statistic for the hypothesis that $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is of rank $r$. Let $L(d, b, r)$ be the profile likelihood function, where $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_k, \boldsymbol{\Omega}$ have been concentrated out by regression and reduced rank regression; see JN (2010b) for details. To calculate the LR statistic, we maximize the profile likelihood function both under the hypothesis $H_r : \mathrm{rank}(\boldsymbol{\Pi}) = r$ and under $H_p : \mathrm{rank}(\boldsymbol{\Pi}) = p$. Thus $\mathrm{LR}(q) = 2\log\big(L(\hat{d}_p, \hat{b}_p, p)/L(\hat{d}_r, \hat{b}_r, r)\big)$, where $q = p - r$,

$$L(\hat{d}_p, \hat{b}_p, p) = \max_{d,b} L(d, b, p), \text{ and } L(\hat{d}_r, \hat{b}_r, r) = \max_{d,b} L(d, b, r).$$

The asymptotic distribution of this test statistic is derived in JN (2010b), in which it is shown that, when $0.5 < b \leq d$,

$$\mathrm{LR}(q) \xrightarrow{d} \mathrm{Tr}\left\{\int_0^1 d\boldsymbol{W}(s)\boldsymbol{W}_b'(s) \left(\int_0^1 \boldsymbol{W}_b(s)\boldsymbol{W}_b'(s)\,ds\right)^{-1} \int_0^1 \boldsymbol{W}_b(s)d\boldsymbol{W}'(s)\right\}, \qquad (2)$$

where the vectors $d\boldsymbol{W}$ and $\boldsymbol{W}_b$ are both of dimension $q = p - r$. The former is the increment of ordinary vector Brownian motion, and the latter is vector fractional Brownian motion of type II defined as

$$\boldsymbol{W}_b = \boldsymbol{\Gamma}(b)^{-1} \int_0^u (u - s)^{b-1} d\boldsymbol{W}. \qquad (3)$$

The asymptotic distribution of the test statistic $\mathrm{LR}(q)$ defined in (2) depends on both $b$ and $q = p - r$. The dependence on the unknown parameter $b$ complicates empirical analysis compared to the cointegrated VAR model. One approach would be to to simulate the asymptotic distribution on a case-by-case basis, but this would be both impractical and imprecise, not to mention time-consuming. As a more practical alternative, we derive NDFs that can be used in practice to obtain accurate asymptotic critical values and $P$ values for any given estimate of $b$ (and a given dimension $q$).

## 3  Numerical distribution functions

In order to simulate $\mathrm{LR}(q)$, we need to generate discrete-time analogs of $\boldsymbol{W}_b$ and $d\boldsymbol{W}$. The latter is easy; we just use standard normal random $q$–vectors $\boldsymbol{\varepsilon}_t$. The former is a little harder. We discretize (3) and use[1]

$$\boldsymbol{z}_t = T^{1/2-b}\boldsymbol{\Gamma}(b)^{-1} \sum_{s=1}^{t-1} (t - s)^{b-1}\boldsymbol{\varepsilon}_s. \qquad (4)$$

---

[1]Alternative simulation strategies are discussed in Davidson and Hashimzade (2009), although they focus on type I processes.

Note that $z_t$ depends on $\varepsilon_s$ for $s < t$ but not on $\varepsilon_t$. Then the simulated test statistic for the null hypothesis $H_r$ against the alternative $H_p$ is

$$\tau_q^b(T) = \text{Tr}\left\{ T^{-1/2} \sum_{t=1}^{T} \varepsilon_t z_t' \left( T^{-1} \sum_{t=1}^{T} z_t z_t' \right)^{-1} T^{-1/2} \sum_{t=1}^{T} z_t \varepsilon_t' \right\}, \tag{5}$$

which depends on $b$, $q$, and $T$.

Since we are interested in the asymptotic distribution of $\text{LR}(q)$, $\tau_q^b(T)$ should ideally be simulated for $T = \infty$. Of course, this is impossible. Even using very large values of $T$ is infeasible, because, for large $T$, the cost of simulation is roughly proportional to the square of $T$. The solution, as in MacKinnon (1996), is to simulate $\tau_q^b(T)$ for a large number of values of $T$, which range from quite small to moderately large, and use the results to estimate the asymptotic distribution by means of response surface regressions on estimated quantiles.

Most of the computational cost is incurred in generating the $\varepsilon_t$ and especially the $z_t$. To reduce this cost, we generate them for $q = 1, 2, \ldots, 12$ simultaneously. For each value of $T$, we perform 200 experiments, and for each experiment we generate 1,199,999 realizations of (5) for $q = 1$, 599,999 realizations for $q = 2$, 399,999 realizations for $q = 3$, 299,999 realizations for $q = 4$, 199,999 realizations for $q = 5$ and $q = 6$, and 99,999 realizations for $7 \leq q \leq 12$. Thus the same random numbers enter into the calculations for all values of $q$. Of course, this scheme implies that our results will be most accurate for $q = 1$ and least accurate for $q \geq 7$.

The numbers of realizations always end in 99,999 so as to make the calculation of quantiles easy. For an experiment based on $B$ realizations, the estimate of the $\pi$ quantile is just number $\pi(B + 1)$ in the list of the $\tau_q^b(T)$ sorted from smallest to largest, provided that $\pi(B + 1)$ is an integer. For each of the 200 experiments for each value of $T$, we store 221 estimated quantiles (.0001, .0002, .0005, .001, \ldots, .010, .015, \ldots, .990, .991, \ldots, .999, .9995, .9998, .9999).

In principle, we are interested in all values of $b$ in the interval $0.5 < b \leq 2$. In practice, however, we must limit our simulations to a finite number of values and use an interpolation scheme. After some experimentation, we chose to perform the simulations for 31 values of $b$: $0.51, 0.55, 0.60, \ldots, 2.0$. The lowest value is 0.51 rather than 0.50 because $b = 0.50$ is not covered by the theory. In principle, we could use any value strictly greater than 0.5, but it seems unwise to get too close.

For each value of $b$, $q$, and $T$, we obtain 200 estimates of the 221 quantiles. We then average these across the 200 experiments to obtain estimated quantiles $\hat{F}_q^b(T, \pi)$, where $\pi$ takes on 221 values, along with standard errors $s_q^b(T, \pi)$. These estimated quantiles are the raw material from which we estimate NDFs by means of 221 response surface regressions for each value of $b$ and $q$. These regressions use functions of $T$ to explain the $\hat{F}_q^b(T, \pi)$.

Even though our experiments were carefully designed to minimize computational cost, they required several years of CPU time. They also used so many random numbers that it was essential to employ a 64-bit random number generator. We actually combined two such generators, which were taken from L'Ecuyer *et al.* (1993).

We initially attempted to estimate response surface regressions similar to those used in MacKinnon (1996) and MacKinnon, Haug, and Michelis (1999). These involve regressing

the $\hat{F}_q^b(T, \pi)$ on a constant and a few powers of $1/T$, using a form of weighted least squares that is described below. However, this does not work well for many values of $b$. Note that the normalizations in $T$ on the two different terms in (5) are $T^{-b}$ and $T^{-2b}$, respectively. Hence, the bias and other approximation and remainder terms that make (5) differ from the right-hand side of (2) are probably (asymptotically) proportional to powers and/or square roots of such terms. Based on this argument, we use regressors that are powers of $T^{-b}$ and/or $T^{-b/2}$, and we found that this generally worked much better than the earlier approach.

We utilize ten response surface models. For every combination of $q$ and $b$, at least one of them fits the data as well as we would expect a true model to fit, perhaps after dropping some observations that correspond to small values of $T$. Moreover, this is true for every model in at least some cases. The models are:

$$\text{Models } 1, 2, 3, 4: \quad 1, \ T^{-b}, \ [T^{-2b}], \ [T^{-3b}], \ [T^{-4b}]$$
$$\text{Models } 5, 6, 7, 8: \quad 1, \ T^{-b/2}, \ [T^{-b}], \ [T^{-3b/2}], \ [T^{-2b}]$$
$$\text{Models } 9, 10: \qquad 1, \ T^{-b/2}, \ T^{-b}, \ T^{-2b}, \ [T^{-3b}]$$

Here square brackets indicate that a regressor does not appear in the lowest-numbered model but does appear in at least one higher-numbered model, with regressors being added sequentially from left to right as the model number increases. Thus Model 1 has just two regressors, namely, a constant and $T^{-b}$, while Model 4 has five regressors. We also experimented with a few models that had more than five regressors. However, we found that these often showed evidence of severe overfitting, and they sometimes even had regressor matrices that were numerically singular.

What we are interested in for all the response surfaces is the constant term, which can be interpreted as an estimate of the $\pi$ quantile of the asymptotic distribution. Ideally, all models that fit well will lead to similar estimates of the constant term. As we explain below, that is true for the larger values of $b$ but not always for the smaller ones.

The estimation method we use is weighted least squares, which can be interpreted as a form of GMM. Following MacKinnon (1996), for every value of $b$, $q$, and $\pi$, we initially regress the square of $s_q^b(T, \pi)$ on a constant, $1/T$, and $1/T^2$ and use the inverse square roots of the fitted values from that regression as weights. We could have skipped the initial regression and simply used $1/s_q^b(T, \pi)$, but the smoothing that results from using fitted values makes the weights much less variable and should produce slightly more efficient estimates.

Provided the weights are asymptotically correct, the sum of squared residuals (SSR) from the weighted regression is asymptotically (in $n$, the number of values of $T$) distributed as $\chi^2(n-k)$, where $k$ is the number of regressors. Thus it is very easy to tell when a response surface regression fits well and when it fits badly. In the former case, the SSR will be less than $n-k$ for more than half the values of $\pi$ and less than the .95 quantile of the $\chi^2(n-k)$ distribution for about 95% of them. In the latter, it will usually be a very big number.

If a particular response surface regression does not fit well, there are potentially two ways to solve the problem. One is to move to a more complicated model (for example, from Model 3 to Model 4) or to a model from a different family (for example, from Model 3 to Model 7). The other is to drop the observation(s) corresponding to one or more of the lowest values of $T$, thus making it easier for the response surface to fit the remaining values. In

the latter case, the sample size becomes $N < n$ instead of $n$. Of course, either moving to a more complicated model or dropping observations necessarily increases the standard error of the estimated constant term. Both effects can be quite large.

In MacKinnon (1996) and other previous work on numerical distribution functions, informal model selection procedures were used to choose the best model and value of $N$ for each set of 221 estimated quantiles. Note that it is essential that these be the same for all $\pi$ for a given set of experiments, since otherwise the resulting NDF might easily be non-smooth or even non-monotonic.

In this paper, however, with far more models to choose among, we discovered that model selection did not always work well, especially for smaller values of $b$. The chosen model, and consequently the NDF, sometimes changed substantially between adjacent values of $b$, even though the relative fits of alternative models changed only modestly. Since we wish to use interpolation for values of $b$ not directly covered by our experiments, it is very important that the NDFs vary smoothly as $b$ changes. Instead of model selection, we therefore use a form of model averaging; see Buckland *et al.* (1997) and Hansen (2007). To our knowledge, this is the first use of model averaging in this context.

We use ten different functional forms for the response surface regressions, namely, Models 1 through 10 described above. Based on experience, we also use 10 different values of $N$, namely, $n$, $n - 1$, ..., $n - 9$. Thus there are 100 combinations of model and sample size. Our criterion function for model $j$ with $N$ observations is

$$Q(j, n, N) = \frac{1}{221} \sum_{\pi} \Big( \mathrm{SSR}_j(\pi) + 2\big(k_j + (n - N)\big) \Big),$$

where $\mathrm{SSR}_j(\pi)$ denotes the SSR of model $j$ for quantile $\pi$, $k_j$ denotes the number of coefficients estimated for model $j$, and the summation over $\pi$ means that we are summing over all 221 quantiles. This criterion function penalizes both extra parameters and reduced numbers of observations symmetrically. The weight given to a particular model/sample combination with $j = J$ and $N = M$ is then

$$W(J, n, M) = \frac{\exp\big(-Q(J, n, M)\big)}{\sum_{j=1}^{10} \sum_{N=n-9}^{n} \exp\big(-Q(j, n, N)\big)}. \tag{6}$$

The formula (6) is based on one suggested by Buckland *et al.* (1997). It gives high weights to models that fit well and are parsimonious, and it ensures that the weights are nonnegative and sum to one. When the weight for any model/sample combination is initially less than $10^{-4}$, it is set to zero, and the remaining weights are rescaled so that they sum to one. This ensures that poorly fitting model/sample combinations have no effect on the results. The final estimate of each of the 221 asymptotic quantiles, for given $b$ and $q$, is then

$$\hat{F}_q^b(\infty, \pi) = \sum_{j=1}^{10} \sum_{N=n-9}^{n} W(j, n, N)\hat{\gamma}_\pi(j, N),$$

where $\hat{\gamma}_\pi(j, N)$ denotes the estimated constant term in the response surface regression for quantile $\pi$ based on model $j$ with $N$ observations, that is, $N$ values of $T$.

We have not yet said anything about $n$, the number of values of $T$, and what those values are. In our initial experiments, we used the following 38 values:

$$50, 55, \ldots, 95, 100, 110, \ldots, 290, 300, 350, \ldots, 550, 600, 700.$$

This choice produced satisfactory results for the largest values of $b$, but not for smaller values. We therefore added some additional, larger values of $T$ for some cases. For $b \geq 1.50$, we use the 38 values given above. For $1.00 \leq b \leq 1.45$, we use 41 values, adding 800, 900, and 1000. For $0.75 \leq b \leq 0.95$, we use 46 values, adding 1100, 1200, 1300, 1400, and 1600 as well. Finally, for $b \leq 0.70$, we use 48 values, adding 1800 and 2000 as well. Each set of 200 experiments for the largest values of $T$ took several days of computer time.

The reason that larger values of $T$ are needed as $b$ becomes smaller is that the shape of the relationship between $\hat{F}_q^b$ and $T$ changes with $b$. Moreover, experimental results for small values of $T$ become less informative about the asymptotic distribution as $b$ becomes smaller. Both of these results are evident in Figure 1, which graphs $\hat{F}_q^b(T, \pi)$ against $T$ for $\pi = 0.95$ and several values of $b$. In the top panel, $q = 1$, and in the bottom panel, $q = 12$. For the larger values of $b$, the relationship appears to be monotonic, but this is not true for $b = .70$ and $b = .51$.

It is clear from Figure 1 that ignoring the presence of the parameter $b$ would lead to substantial errors in inference, since quantiles, and therefore critical values and $P$ values, depend in a non-trivial way on $b$. It would therefore be impractical to use tables of simulated critical values, which would extend to many pages. Moreover, such tables would provide less information than the $P$ values provided by our numerical distribution functions; see the next section.

From Figure 1, it is evident that our response surface regressions are likely to have a much easier time fitting the data for large values of $b$ than for small ones. In the former case, all response surfaces that fit well tend to yield very similar estimates of $F_q^b(\infty, \pi)$. Thus, for large $b$, our NDFs suffer from very little model uncertainty. In the latter case, however, several response surfaces that all fit well can yield somewhat different estimates of $F_q^b(\infty, \pi)$. In consequence, despite the very large number of simulations that we performed, the NDFs for small values of $b$ may suffer from more model uncertainty.

To make the above remarks more concrete, consider the .95 quantile (that is, the .05 critical value) for $q = 1$. When $b = 1.50$, there are 18 model/sample combinations with weights greater than 0.02. Among these, the smallest estimate of $F_1^{1.50}(\infty, 0.95)$ is 4.54955, the largest is 4.55165, and the weighted average is 4.55069. In contrast, when $b = 0.51$, there are 16 model/sample combinations with weights greater than 0.02. Among these, the smallest estimate of $F_1^{0.51}(\infty, 0.95)$ is 3.62530, the largest is 3.66211, and the weighted average is 3.64518. In the former case, all the response surface regressions yield what is for all practical purposes the same estimate, and the weighted average is almost certainly very accurate. In the latter case, they yield estimates that vary somewhat. With the model averaging approach, we avoid making an arbitrary choice among the model/sample combinations that all fit quite well.

In any case, it does not make sense to worry much about the possible inaccuracy of our NDFs for small values of $b$. As can be seen from the above example (the case $b = 0.51$ is the most extreme one) and from Figure 1, any such inaccuracy is small relative to the errors

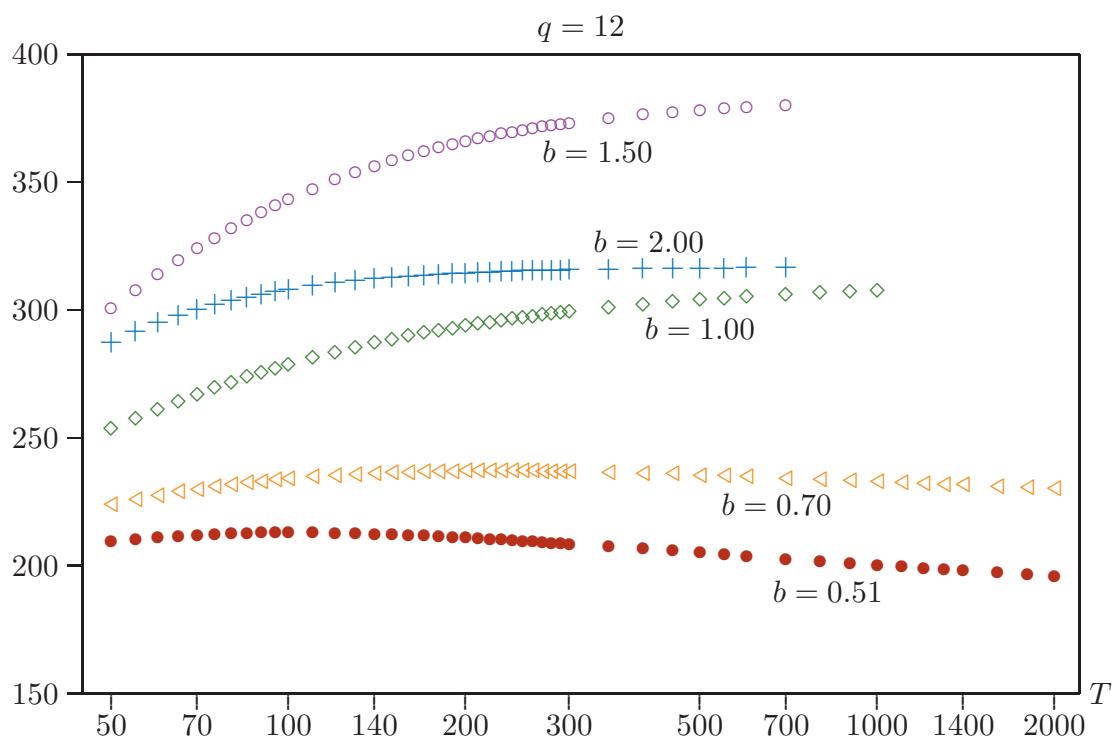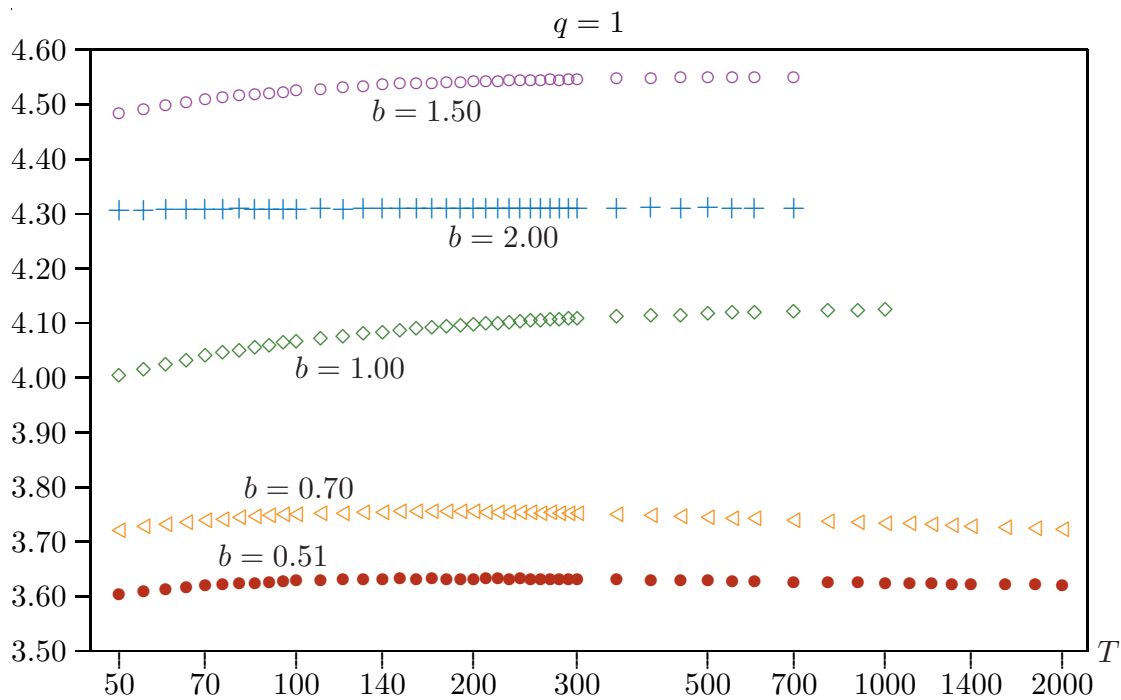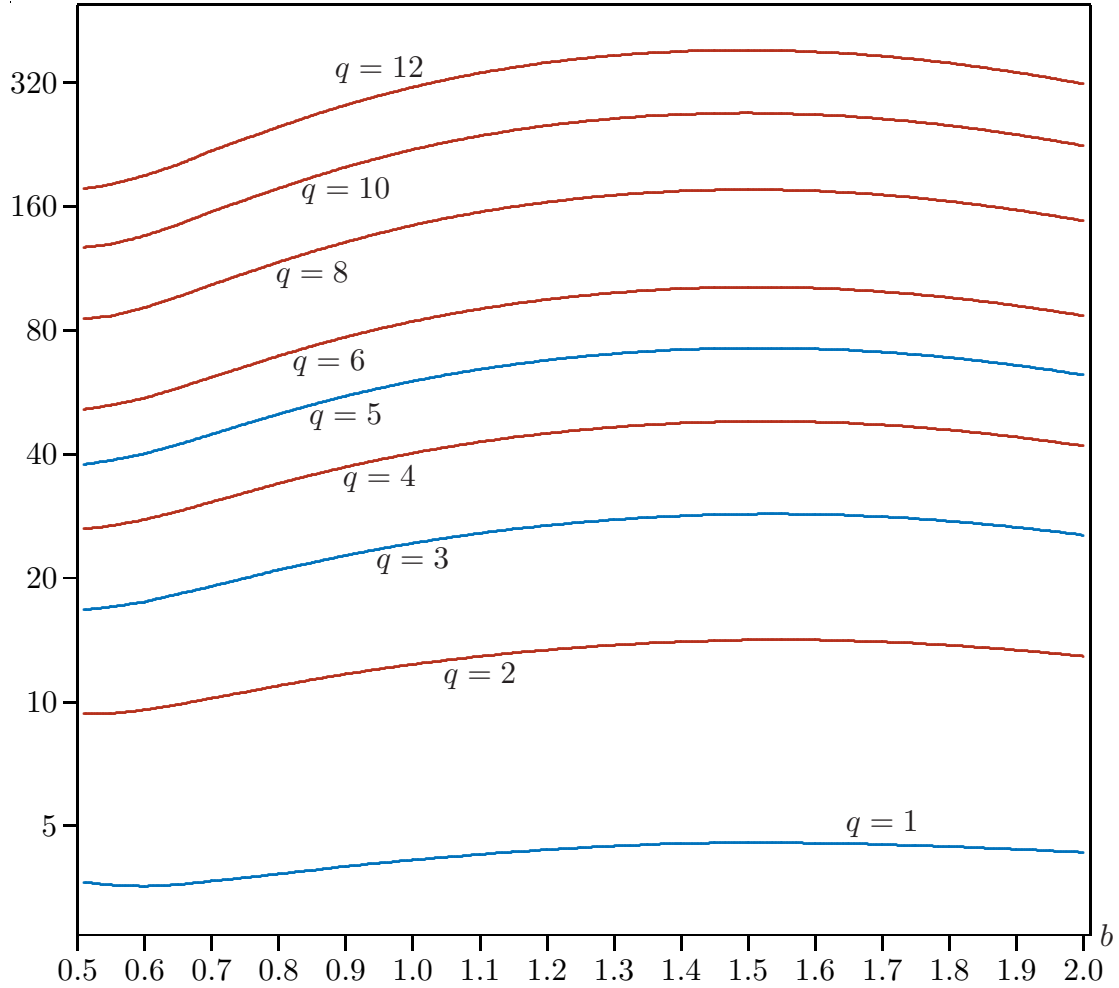Figure 1: Estimated .95 quantiles as functions of $T$

Figure 2: Estimated .95 asymptotic quantiles as functions of $b$

that must inevitably arise from using asymptotic rather than finite-sample distributions and from using estimates of $b$ rather than true values.

It is interesting that the curves graphed in Figure 1 appear to be very smooth, even though these are simply raw results that have not been smoothed in any way. The curves appear to be smooth because they are based on a great many simulations, and there is consequently very little experimental error. It is because of this that poorly specified response surfaces produce very large values of $\mathrm{SSR}_j(\pi)$.

From Figure 1, it is evident that the asymptotic critical values are not monotonic in $b$. The relationship between $b$ and the asymptotic .95 quantiles (.05 critical values) is shown in Figure 2 (on a logarithmic scale) for all even values of $q$ as well as for $q = 1, 3, 5$. For all values of $q$, the critical value initially rises slowly as $b$ increases (it actually drops slightly at first for $q = 1$), then rises quite rapidly for a while, achieves a maximum around $b = 1.5$, and declines quite substantially thereafter.

Figure 2 contains all the information that our experiments provide about asymptotic .05

critical values (.95 quantiles), except for $q = 7, 9, 11$, which were omitted to avoid cluttering the figure. In principle, one could simply read a desired critical value from the figure. However, figures of this sort are not very convenient for investigators who wish to calculate critical values or $P$ values. In the next section, we discuss some computer programs which accompany this paper and make it easy to do so.

## 4  Critical values and $P$ values

The first step in estimating critical values and $P$ values for a specified value $\bar{b}$ is to calculate $\hat{F}_q^{\bar{b}}(\infty, \pi)$ using the $\hat{F}_q^b(\infty, \pi)$ for $b$ close to $\bar{b}$. As Figure 2 makes clear, the asymptotic distributions are smooth functions of $b$, so it should be possible to do this in a variety of ways. We use the weighted local quadratic regression:

$$w(i)\hat{F}_q^{b(i)}(\infty, \pi) = \eta_1 w(i) + \eta_2 w(i)b(i) + \eta_3 w(i)b^2(i) + u(i), \tag{7}$$

where $i$ runs from 1 to 31, $b(1) = 0.51$, $b(2) = 0.55$, and so on up to $b(31) = 2.00$, and $w(i) = \max(0, 1 - 5|b(i) - \bar{b}|)$. With this weighting scheme, only values of $b$ near $\bar{b}$ receive positive weights. For example, if $\bar{b} = 0.72$, there are eight such values:

$$.55(.15), .60(.40), .65(.65), .70(.90), .75(.85), .80(.60), .85(.35), .90(.10).$$

Here the first number is $b(i)$ and the second is the weight $w(i)$. This case is typical, in that eight values of $b$ receive positive weights. However, there may be as few as five when $\bar{b}$ is close to 0.51 or 2.00. Our estimate $\hat{F}_q^{\bar{b}}(\infty, \pi)$ is simply $\hat{\eta}_1 + \hat{\eta}_2\bar{b} + \hat{\eta}_3\bar{b}^2$, where the $\hat{\eta}_j$ are the estimates from regression (7).

In order to obtain a $P$ value for any observed test statistic $\text{LR}(q)$, a procedure for interpolating between the 221 values of $\hat{F}_q^{\bar{b}}(\infty, \pi)$ is needed. The procedure we use is a modified version of one proposed in MacKinnon (1996). Consider the cubic regression

$$G(\pi) = \gamma_0 + \gamma_1 \hat{F}(\pi) + \gamma_2 \hat{F}^2(\pi) + \gamma_3 \hat{F}^3(\pi) + e_\pi, \tag{8}$$

where $\hat{F}(\pi)$ denotes $\hat{F}_q^{\bar{b}}(\infty, \pi)$, and $G(\cdot)$ is an inverse CDF. The left-hand side variable is transformed in order to make the relationship between it and $\hat{F}(\pi)$ closer to linear. In MacKinnon (1996), which dealt with Dickey-Fuller tests, $G(\cdot)$ was the inverse cumulative standard normal CDF. That would be a bad choice here, because $F(\cdot)$ in no way resembles the inverse standard normal distribution. Instead, we note that if standard asymptotics applied, then $\text{LR}(q) \xrightarrow{d} \chi^2(q^2)$. We therefore define $G(\pi)$ to be the inverse of the $\chi^2(q^2)$ CDF evaluated at $\pi$. The precise choice of $G(\cdot)$ does not matter very much.

It may seem a bit odd that the regressors in (8) are stochastic and the regressand is not. However, because the estimated quantiles are generally very accurate, the errors in variables bias that this induces is very small. This regression is estimated using only a small number of points, generally 9, in the neighborhood of the observed test statistic $\text{LR}(q)$. We first find the value of $\hat{F}(\pi)$ that is closest to $\text{LR}(q)$ and use the data for that value of $\pi$ and the four values on either side of it. Slightly different procedures are used if the initial value of $\pi$ is less than .002 or greater than .998. The estimated $P$ value is simply

$$\hat{P} = 1 - G^{-1}\left(\hat{\gamma}_0 + \hat{\gamma}_1 \text{LR}(q) + \hat{\gamma}_2 \text{LR}(q)^2 + \hat{\gamma}_3 \text{LR}(q)^3\right),$$

where the $\hat{\gamma}_i$ are the parameter estimates.

The $\hat{F}_q^{\bar{b}}(\infty, \pi)$ already give us asymptotic critical values for tests at any conventional significance level. However, we can obtain ones for tests at any level at all that should be slightly more accurate by turning regression (8) around:

$$\hat{F}(\pi) = \delta_0 + \delta_1 G(\pi) + \delta_2 G^2(\pi) + \delta_3 G^3(\pi) + e'_\pi. \tag{9}$$

Of course, this is not actually the inverse of equation (8), but it involves the same sort of approximation. As before, regression (9) is (usually) estimated for 9 observations around the desired level of the test. The estimated critical value for a test at level $\alpha$ is then simply

$$\hat{\delta}_0 + \hat{\delta}_1 G(1 - \alpha) + \hat{\delta}_2 \big(G(1 - \alpha)\big)^2 + \hat{\delta}_3 \big(G(1 - \alpha)\big)^3,$$

where the $\hat{\delta}_i$ are the parameter estimates.

The $\hat{F}_q^b(\infty, \pi)$ and a Fortran program that uses regressions (7), (8), and (9) to obtain estimated asymptotic $P$ values and critical values from them are available from the authors' websites. The program is intended to be run interactively, but everything of any substance is done in subroutines, so that the essential parts can easily be extracted and incorporated into larger programs.

## 5   An Empirical Example

In this section, we illustrate our methods and our program with a brief empirical application. The data are monthly Gallup opinion poll data on support for the Conservative and Labour parties in the United Kingdom. They cover the period from January 1951 to November 2000, for a sample of $T = 599$ months. A shorter version of this dataset was used in Byers *et al.* (1997). The two series have been logistically transformed, so that, if $y_t$ denotes an observation on the original series, it is mapped into $x_t = \log(y_t/(100 - y_t))$.

Using an aggregation argument and a model of voter behavior, Byers *et al.* (1997, 2002) show that aggregate opinion poll data may be best modeled using fractional integration methods and obtains values of the integration parameter $d$ in the range of 0.7–0.8. Dolado *et al.* (2002) also concludes that a fractional unit root is present with very similar estimated values of $d$. The same model of voting behavior and similar conclusions were derived independently using U.S. data in Box-Steffensmeier and Smith (1996). The analysis was extended to fractional cointegration in Davidson *et al.* (2006).

We first conduct a univariate analysis of each of the two series. We estimate only the model with $d = b$ because it allows $k = 0$,[2] which is relevant for these data. For example, Byers *et al.* (1997) shows that ARFIMA(0,$d$,0) models appear to fit the data well.

In the analysis of nonstationary autoregressive processes, it is necessary to condition on initial values. Specifically, it is usually assumed that initial values are observed but not modeled and inference is conditional on them. For example, in the nonstationary AR($k$) model, inference is conditional on the first $k$ observations. In principle, infinitely many initial values are needed in fractional models, but in practice truncation is applied; see the discussion in JN (2010a, p. 54). We chose to condition on the first 60 observations, i.e., five years of data. Results for other choices (eight and ten years) are very similar.

---

[2]In the model with both $d$ and $b$ freely varying, the parameter $b$ is only identified if $k \geq 1$.

Table 1: Opinion poll data: univariate results

| $r$ | $k$ | $\hat{d}\,(=\hat{b})$ | s.e.$(\hat{d})$ | Log $L$ | LR$(\mu)$ | LR$(d)$ | LR$(q)$ | $q$ | $P$ value | 10% cv. | 5% cv. | 1% cv. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Conservative Party | | | | | | | | | | | | |
| 0 | 0 | 0.7657 | 0.0381 | 405.9043 | 1.2540 | 38.1400 | 0.0616 | 1 | 0.8002 | 2.6453 | 3.7639 | 6.5279 |
| 1 | 0 | 0.7610 | 0.0379 | 405.9351 | 2.8188 | 33.4510 | – | – | – | – | – | – |
| Labour Party | | | | | | | | | | | | |
| 0 | 1 | 0.8092 | 0.0595 | 362.5316 | 0.0034 | 14.2686 | 1.5317 | 1 | 0.2176 | 2.7095 | 3.8341 | 6.6088 |
| 1 | 1 | 0.8428 | 0.0561 | 363.2974 | 2.2254 | 8.5268 | – | – | – | – | – | – |

Note: Results from univariate estimation of model (1) are presented for the logistically transformed opinion poll data, which are monthly from 1951.1 to 2000.11. We use the first 60 observations as initial values and estimate the model using the remaining $T = 539$ observations.

Results from univariate estimation of model (1) appear in Table 1. For each of the two series, we estimate the model with $r = 0$, i.e. $\mathbf{\Pi} = 0$, and with $r = 1$. The lag length $k$, indicated in the second column, was chosen by a general-to-specific procedure where lags corresponding to insignificant lag coefficients (as measured by LR tests) were dropped sequentially until the highest order lag had a significant estimated coefficient while tests for serial correlation in the residuals remained insignificant. The likelihood ratio test statistic against the inclusion of an unrestricted constant term in the model is reported in the table as LR$(\mu)$. Although the theory for this test has not yet been developed, we conjecture based on the corresponding results for the cointegrated VAR model that LR$(\mu) \overset{d}{\to} \chi^2(p)$. Based on this test, there are no indications that a constant term is needed for either series.

We also fitted standard augmented Dickey-Fuller models to these data series. In that case, the number of lagged differences, $k$, required to remove evidence of serial correlation in the residuals is considerably larger. We take this as strong evidence that the true model is one of fractional integration, since several more lagged differences are required by the I(1) model to attempt to fit this feature of the data. As further evidence, we report LR tests of the I(1) model against the fractional model under the heading LR$(d)$. Since LR$(d) \overset{d}{\to} \chi^2(1)$ under the null that $d = 1$, these tests reject very strongly.

The test statistic LR$(q)$ (with $q = 1$) reported in the table is the LR test of the hypothesis that $r = 0$ against the alternative that $r = 1$. In the univariate case, this corresponds to testing $\mathbf{\Pi} = 0$ against $\mathbf{\Pi} \neq 0$. Under the null hypothesis, the series is (fractionally) integrated of order $d$, and under the alternative it is integrated of order $d - b = 0$. In the last four columns of the table, we report the output from our program after interactively entering $q\,(= 1)$, $\hat{d}\,(= \hat{b})$, and LR$(q)$. We report the resulting $P$ value of the test and also the 10%, 5%, and 1% critical values.[3] For both series, we are unable to reject the null of a fractional unit root, with $P$ values of 0.8002 and 0.2176, respectively. Thus, the results very clearly indicate that both series are (fractionally) integrated of order $d$.

The estimated fractional integration parameters $\hat{d}$ for each of the two series are very close in value and appear insignificantly different. This result was also found by Davidson

---

[3]Without our program, we would have had to simulate critical values and/or $P$ values conditional on the estimated value $\hat{d}\,(= \hat{b})$. This would have been very time-consuming.

Table 2: Opinion poll data: multivariate FVAR results

| $r$ | $k$ | $\hat{d}\,(=\hat{b})$ | $s.e.(\hat{d})$ | Log $L$ | LR($\mu$) | LR($d$) | LR($q$) | $q$ | $P$ value | 10% cv. | 5% cv. | 1% cv. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0.8147 | 0.0458 | 837.2784 | 1.0886 | 25.5036 | 2.8689 | 2 | 0.7259 | 9.2361 | 11.0494 | 15.0159 |
| 1 | 1 | 0.8438 | 0.0427 | 838.7093 | 3.3080 | 16.3866 | 0.0070 | 1 | 0.9375 | 2.7606 | 3.8907 | 6.6728 |

Note: Results from bivariate estimation of model (1) are presented for the logistically transformed opinion poll data, which are monthly from 1951.1 to 2000.11. We use the first 60 observations as initial values and estimate the model using the remaining $T = 539$ observations.

*et al.* (2006). This opens up the possibility of multivariate modeling of the two series with a common fractional integration order and possibly cointegration. We therefore estimate the bivariate and possibly cointegrated FVAR model (1). The results are presented in Table 2 for $r = 0$ and $r = 1$. Results for $r = 2$ are not presented, because under that hypothesis neither series would be fractionally integrated, and we have already established in Table 1 that they both are. The only result for $r = 2$ that we need is that Log $L = 838.7128$.

Once again, only $k = 1$ lag is needed in the fractional model. In contrast, when estimating a standard cointegrated VAR model for these data, several more lags are needed to yield insignificant serial correlation tests on the residuals. Also, as before, the test statistics LR($d$) of the I(1) model against the fractional model reject very strongly. We take these results as further evidence of the appropriateness of the fractional model for this data set.

In Table 2, we first consider the hypothesis that the rank is $r = 0$, i.e., that there is no cointegration among the two series. The test statistic LR($\mu$) gives no indication that a constant term is needed. We obtain an LR rank test statistic of 2.8689. Using our interactive program with $q = 2$, $\hat{d}\,(=\hat{b}) = 0.8147$, and LR(2) $= 2.8689$, we obtain a $P$ value for the test of 0.7259. The table also includes the 10%, 5%, and 1% critical values for the test as reported by our program. Since we do not reject the null that there is no cointegration among the two (fractionally) integrated opinion poll series, there appear to be two stochastic trends underlying the system. For completeness, the table also shows the results of testing the hypothesis that $r = 1$. If the results of the initial test had been different, this would have been an interesting thing to do.

## 6   Conclusion

We obtain asymptotic numerical distribution functions of likelihood ratio tests for fractional unit roots and cointegration rank. These can easily be used to obtain critical values and $P$ values via a program that we make available. Because these distributions depend on a real-valued parameter, $b$, computational cost is very high. Moreover, the choice of model specification for the response surface regressions used to obtain the NDFs is more involved than is usual. We use model averaging rather than model selection to deal with uncertainty about the correct form of the response surface regressions. The use of our program is illustrated by means of an empirical example that involves opinion poll data, a case in which both theory and empirical results strongly support the use of fractional VAR models.

# 7 References

Box-Steffensmeier, J. M., and Smith, R. M. (1996). The dynamics of aggregate partisanship. *American Political Science Review* **90**, 567–580.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.

Byers, J. D., Davidson, J., and Peel, D. (1997). Modelling political popularity: An analysis of long-range dependence in opinion poll series. *Journal of the Royal Statistical Society, Series A* **160**, 471–490.

Byers, J. D., Davidson, J., and Peel, D. (2002). Modelling political popularity: A correction. *Journal of the Royal Statistical Society, Series A* **165**, 187–189.

Davidson, J. E. H., Peel, D. A., and Byers, J. D. (2006). Support for governments and leaders: Fractional cointegration analysis of poll evidence from the UK, 1960-2004. *Studies in Nonlinear Dynamics & Econometrics* **10**, issue 1, article 3.

Davidson, J., and Hashimzade, N. (2009). Type I and type II fractional Brownian motions: A reconsideration. *Computational Statistics & Data Analysis* **53**, 2089–2106.

Dolado, J. J., Gonzalo, J., and Mayoral, L. (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica* **70**, 1963–2006.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica* **75**, 1175–1189.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.

Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* **24**, 651–676.

Johansen, S. and Nielsen, M. Ø. (2010a). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* **158**, 51–66.

Johansen, S. and Nielsen, M. Ø. (2010b). Likelihood inference for a fractionally cointegrated vector autoregressive model. QED working paper 1237, Queen's University.

L'Ecuyer, P., Blouin, F., and Couture, R. (1993). A search for good multiple recursive random number generators. *ACM Transactions on Modeling and Computer Simulation* **3**, 87–98.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* **11**, 601–618.

MacKinnon, J. G., Haug, A. A., and Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics* **14**, 563–577.

# Research Papers
# 2010

2010-45:     Rasmus Tangsgaard Varneskov and Valeri Voev: The Role of Realized Ex-post Covariance Measures and Dynamic Model Choice on the Quality of Covariance Forecasts

2010-46:     Christian Bach and Stig Vinther Møller: Habit-based Asset Pricing with Limited Participation Consumption

2010-47:     Christian M. Dahl, Hans Christian Kongsted and Anders Sørensen: ICT and Productivity Growth in the 1990's: Panel Data Evidence on Europe

2010-48:     Christian M. Dahl and Emma M. Iglesias: Asymptotic normality of the QMLE in the level-effect ARCH model

2010-49:     Christian D. Dick, Maik Schmeling and Andreas Schrimpf: Macro Expectations, Aggregate Uncertainty, and Expected Term Premia

2010-50:     Bent Jesper Christensen and Petra Posedel: The Risk-Return Tradeoff and Leverage Effect in a Stochastic Volatility-in-Mean Model

2010-51:     Christos Ntantamis: A Duration Hidden Markov Model for the Identification of Regimes in Stock Market Returns

2010-52:     Christos Ntantamis: Detecting Structural Breaks using Hidden Markov Models

2010-53:     Christos Ntantamis: Detecting Housing Submarkets using Unsupervised Learning of Finite Mixture Models

2010-54:     Stefan Holst Bache: Minimax Regression Quantiles

2010-55:     Nektarios Aslanidis and Charlotte Christiansen: Sign and Quantiles of the Realized Stock-Bond Correlation

2010-56:     Anders Bredahl Kock: Oracle Efficient Variable Selection in Random and Fixed Effects Panel Data Models

2010-57:     Charlotte Christiansen, Juanna Schröter Joensen and Jesper Rangvid: The Effects of Marriage and Divorce on Financial Investments: Learning to Love or Hate Risk?

2010-58:     Charlotte Christiansen, Maik Schmeling and Andreas Schrimpf: A Comprehensive Look at Financial Volatility Prediction by Economic Variables

2010-59:     James G. MacKinnon and Morten Ørregaard Nielsen: Numerical distribution functions of fractional unit root and cointegration tests