LUND UNIVERSITY

Bandspectrum Cointegration

Andersson, Fredrik N G

2008

Document Version:
Other version

[Link to publication](#)

Citation for published version (APA):
Andersson, F. N. G. (2008). *Bandspectrum Cointegration*.

Total number of authors:
1

Bandspectrum Cointegration

Fredrik NG Andersson[*]
Department of Economics, Lund University

December 2, 2008

Abstract

Economic theory commonly distinguishes between different time horizons such as the short run and the long run, each with its own relationships and its own dynamics. Engle (1974) proposed a bandspectrum regression to estimate such models. This paper proposes a new estimator for non-stationary panel data models, a bandspectrum cointegration estimator. The bandspectrum cointegration estimator uses first differenced data to avoid spurious results. Such estimates are, however, less efficient than estimates from a model with non-stationary data. Still, simulation results in the paper show that the bandspectrum cointegration estimator is more efficient than common time domain estimators, for example VECM and OLS levels estimators, if the data generating process contains more than one time horizon. The BSCE furthermore identifies all horizons in the data generating process and estimates an individual parameter vector for each, a property that neither time domain estimator possesses.

JEL Classification: C14, C15, C23
Keywords: Cointegration; Bandspectrum Regression; Simulations; Wavelets; Frequency domain

1. Introduction

It is well understood in economics that relationships between macroeconomic variables may vary across time horizons (see for example King and Watson 1996, and Ramsey and Lampart

1998). Englund *et. al.* (1992) suggested that macroeconomic variables should be studied at three horizons – short run shocks, medium run business cycle fluctuations and the long run equilibrium. Friedman and Kuznets (1954) make a similar decomposition of the economy into a temporary, quasi-permanent and permanent component. Understanding the economy at different horizons is important for both policy makers and the general public. For example, policies for the long run can cause short run disturbances in the economy, but by also considering the short run effects the policy maker can take appropriate actions to limit these temporary fluctuations. Furthermore, taking different dynamics into account can yield better forecast models and help the individual to smooth consumption over time (Ramsey 2002 and Crowley 2007).

Engle (1974) proposed a bandspectrum regression for models with many time horizons. The bandspectrum regression is estimated in the frequency domain, because in the frequency domain the time horizons are associated with a particular frequency or band of frequencies. For example, the long run is represented by low frequencies around the zero frequency, while the short run is represented by higher frequencies. Instead of using all frequencies to estimate one model, the data is split into many frequency components and a separate model with different variables and different parameter values, is estimated for each horizon[1].

Many economic time series are non-stationary in the sense that they do not have a finite covariance matrix (Granger 1966, 1980). Regression models with non-stationary data can yield spurious results because standard testing procedures cannot be applied to them (see Yule 1926 and Granger and Newbold 1974). A growing literature on how to estimate such models has thus emerged (see for example Engle and Granger 1987 and Johansen 1988, 1991). The long run information dominates the data generating process for a non-stationary time series (Granger and Joyeux 1980), and regression models with such data can therefore be interpreted as models of the long run. A sample with observations from just one individual, however, may contain too few long run observations to yield efficient parameter estimates. The cointegration literature has, for this and other reasons, widened to also consider panel data models. If it is possible to use the same model (with the same parameter values) for all individuals, pooling individuals increases the number of long run observations, and more efficient parameter estimates are therefore obtained.

The bandspectrum regression that Engle proposed only considers models with stationary data, while Phillips (1991) and Robinson and Marinucci (1998) use the frequency domain to

---

[1] It is naturally possible that some variables are included in both the short run and the long run model.

study non-stationary time series. However, Robinson and Marinucci only use the frequency domain to obtain more efficient parameter estimates of the long-run, and Phillips' estimator only has two time horizons, the long run and the short run. This paper therefore considers a new estimator, the Bandspectrum Cointegration Estimator (BSCE), which allows the estimation of models with many time horizons. The difference compared to the bandspectrum regression, proposed by Engle (1974), is that it is uses non-stationary panel data.

The BSCE has three important properties; (i) it uses first differenced data to avoid spurious regressions, (ii) it pools individuals in a panel to increase the number of long run observations, and (iii) it uses the properties of the frequency domain to estimate a separate model for each time horizon.

Estimating models with first differenced data is usually avoided for two reasons. Firstly, empirical experience shows that it is difficult to obtain parameter estimates of the long run using first differenced data. Secondly, an estimator that uses first differenced data is less efficient than estimators that use the non-stationary levels data (Davidson and MacKinnon 2004). We therefore study the small sample properties of the BSCE in a simulation study, and compare it to three common time domain estimators; the OLS levels estimator, OLS first difference estimator, and a Vector Error Correction Model (VECM). We also compare it to one frequency domain estimator; the Narrow Band Frequency Domain Least Squares (NBFDLS). Two properties are evaluated; the efficiency of the BSCE compared to the other estimators, and whether the BSCE identifies all the horizons in the data generating process.

The simulation model divides the data generating process into three horizons – short run shocks, medium run business cycles and a long run cointegration relationship. The simulation results show that the BSCE accurately identifies, and estimates a separate parameter vector, for all three horizons; it is in fact the only estimator that identifies all horizons. Furthermore, the BSCE estimates of the cointegration vector are less biased and more efficient than the OLS level estimator, OLS first difference, the VECM and NBFDLS. The OLS and VECM estimators are however slightly more efficient than the BSCE when there is one common parameter vector (one horizon) for all frequencies.

The rest of the paper is organized as follows; Section 2 presents the Bandspectrum Regression, Section 3 introduces the Bandspectrum Cointegration Estimator, Section 4 discusses Fourier and wavelet transform, Section 5 contains the simulation results and Section 6 concludes the paper.

2. Bandspectrum Regression

Following, in spirit, Engle (1974), let x_t be an infinite discrete time series with $t \in \{0, \pm 1, \pm 2, ...\}$. The Fourier transform of x_t is

$$F_x(f) = \sum_{t=-\infty}^{\infty} x_t e^{-i2\pi ft},$$

(1)

where the function $F_x(f)$ is the Fourier transform pair, and $-\frac{1}{2} \leq f \leq \frac{1}{2}$. The inverse transform is

$$x_t = \int_{-1/2}^{1/2} F_x(f) e^{i2\pi ft} df.$$

(2)

Let us first consider the regression model where there is only one horizon,

$$y_t = x_t \beta + \varepsilon_t,$$

(3)

where ε_t is white noise with mean zero and variance σ_ε^2. The OLS estimator of β is

$$\hat{\beta} = (x^T x)^{-1} x^T y,$$

(4)

Parcival's theorem states that

$$\sum_{t=-\infty}^{\infty} |x|^2 = \int_{-1/2}^{1/2} F_x(f) \overline{F_x(f)} df,$$

(5)

where $\overline{F_x(f)}$ is the complex conjugate of $F_x(f)$ (Percival and Walden 2006). For a real valued function we can therefore write

$$\sum_{t=-\infty}^{\infty} |x|^2 = \sum_{t=-\infty}^{\infty} x^2 = (x^T x) = \int_{-1/2}^{1/2} F_x(f) \overline{F_x(f)} df.$$

(6)

It naturally follows that

$$x^T y = \int_{-1/2}^{1/2} F_x(f) \overline{F_y(f)} df,$$

(7)

and the OLS estimator, (2.4), can thus be written

$$\hat{\beta} = (x^T x)^{-1} x^T y = \left(\int_{-1/2}^{1/2} F_x(f) \overline{F_x(f)} df \right)^{-1} \int_{-1/2}^{1/2} F_x(f) \overline{F_y(f)} df.$$

(8)

Equation (2.8) can be rewritten as

$$\hat{\beta} = (x^T x)^{-1} x^T y = \left(\int_{-1/2}^{1/2} s_x(f) df \right)^{-1} \int_{-1/2}^{1/2} s_{xy}(f) df,$$

(9)

where

$$s_x(f) = F_x(f) \overline{F_x(f)},$$

(10)

is the spectral density function for x, and

$$s_{xy}(f) = F_x(f)\overline{F_y(f)}, \tag{11}$$

is the cross spectral density function for x and y. The cross spectral density for x and y is[2,3]

$$s_{xy}(f) = \beta s_x(f), \tag{12}$$

(Brockwell and Davis, 1991). The OLS estimator can thus be written as

$$\hat{\beta} = \frac{\beta \int_{-1/2}^{1/2} s_x(f) df}{\int_{-1/2}^{1/2} s_x(f) df}. \tag{13}$$

If x is independent of ε the Gauss-Markow theorem implies that the OLS estimator is the best linear unbiased estimator (BLUE) of β.

Economic theory commonly suggest that the economy should be divided into different time horizons such as the short run, the medium run and the long run each with its own relationships and dynamics. The simple regression model, which only considers one horizon, is therefore inadequate. In the frequency domain each horizon is represented by its own frequency or band of frequencies. For example, the long run is represented by the low frequencies around the zero frequency and the short run by the highest frequency close to $|f| = \frac{1}{2}$. A more appropriate regression model with many time horizons is thus

$$y_t^f = x_t^f \beta_f + \varepsilon_t^f \tag{14}$$

for $f \in [-1/2, 1/2]$ where[4]

$$
\begin{aligned}
y_t^f &= F_y(f)e^{i2\pi ft} + F_y(-f)e^{i2\pi ft} \\
x_t^f &= F_x(f)e^{i2\pi ft} + F_x(-f)e^{i2\pi ft} \\
\varepsilon_t^f &= F_\varepsilon(f)e^{i2\pi ft} + F_\varepsilon(-f)e^{i2\pi ft}
\end{aligned}
\tag{15}
$$

It is possible that $\beta_f = \beta_{f'}$ for some $|f|, |f'| \in [0, 1/2]$, this means that the time horizon is represented not by an individual frequency but by a wider band of frequencies.

---

[2] Assuming that x and ε are orthogonal, and that we have infinite many observations such that $s_{x\varepsilon}(f) = 0$. Otherwise $s_{yx}(f) = \beta s_x(f) + s_{x\varepsilon}(f)$

[3] Equation (12) assumes that the data generating process is given by Equation (2.3). Let $y_t = \sum_{j=-\infty}^{\infty} \beta_j x_{t-j}$, in this case we have that; $s_y(f) = \sum_{j=-\infty}^{\infty} \beta_j e^{-i2\pi f j} s_x(f)$.

[4] The negative frequencies are a mirroring of the positive frequencies for a real valued process (see Percival and Walden 2006).

Consider what happens when the estimator (8) is used on the model (14). We now get

$$\hat{\beta} = \int\limits_{-1/2}^{1/2} \frac{\beta_f s_x(f)}{\left(x^T x\right)} df \tag{16}$$

The OLS estimate is therefore a weighted average of the individual parameter vectors for the different frequencies, where the weights depend of the relative importance of each frequency in the data generating process of x.

3. Bandspectrum Cointegration

A common class of time series in economics are time series that do not have a finite covariance matrix, so called non-stationary time series (see Granger 1966 and 1980). Regression models with non-stationary data can cause spurious results because standard testing procedures cannot be applied to them.

Let x_t be an I(d) process where d is the integration order, υ_t is I(0) and iid $\left(0, \sigma_\upsilon^2\right)$. The data generating process for x_t is given by

$$(1-Z)^d x_t = \upsilon_t \tag{17}$$

where Z is the back shift operator, (Granger and Joyeux 1980), and the spectral density function for x is

$$s_x(f) = \frac{\sigma_\upsilon^2}{\left(4\sin^2(\pi f)\right)^d} \tag{18}$$

(see Brockwell and Davis 1991). The auto-covariance function, $E[x_t x_{t+h}]$, is given by

$$\gamma(h) = \int\limits_{-1/2}^{1/2} s_x(f) e^{i2\pi f h} df \tag{19}$$

Note that x_t does not have a finite covariance matrix if $|d| \geq 0.5$. It is the zero frequency which is causing the non-stationarity, as is evident from (18) and (19); all other frequencies have a finite covariance matrix. This result has important implications for the OLS estimator in (8). The zero frequency component dominates all other frequency components if $d \geq 0.5$. If the data generating process is given by (14) the estimated parameter vector converges in probability to β_0 which represents the long run cointegration relationship. An OLS estimator using non-stationary data is therefore an estimator of the long run.

Taking the first difference of a fractionally differenced time series, as in equation (17), reduces the integration order by one. The first difference of x is hence stationary if x is integrated I(d) with $0.5 \leq d < 1.5$. Problems such as spurious regressions that emerge with

non-stationary time series can therefore be avoided by using first differenced data instead of levels. There are, however, two disadvantages with using first differenced data. Firstly, such estimates are less efficient than estimates using non-stationary levels data, and secondly empirical experience show that it is difficult to estimate long run relationships using first differenced data.

The second problem is probably caused by a true data generating process with many time horizons, i.e. a data generating process given by (14) and not (3). It is common for the short run to be represented by a wider band of frequencies in the frequency domain than the long run. For example assume that the data generating process contains two horizons; business cycle fluctuations and the long run. The business cycle lasts for say four years and the sample consists of yearly data. In this case the business cycle would be represented by, approximately, the frequencies 1/8 to 1/2 and the long run by the frequencies 0 to 1/8. The low frequencies around the zero frequency do not dominate the data generating process for a stationary time series. For an I(0) process, for example, the spectrum is flat. Because the business cycle is represented by the greater part of the spectrum (75% of all frequencies) these will have a larger effect on the OLS estimate than the low frequencies.

Some cointegration models in the time domain, for example, Vector Error Correction models (VECM)[5], mix levels and first differences to estimate the short run and long run simultaneously. The long run is estimated when the data is in levels while the short run is estimated using first differenced data. This only works, however, for two horizons, and only when the short run is represented by a relatively wider band of frequencies than the long run.

Engle (1974) proposed a bandspectrum regression for models with many horizons that is estimated in the frequency domain. Each horizon is relatively easily identified in the frequency domain, because they are represented by their own set of frequencies, and it is therefore straightforward to estimate a separate model for each horizon in this domain. Phillips (1991) and Robinson and Marinucci (1998) have studied frequency based estimators of the cointegration vector. However they do not fully explore the properties of the frequency domain to model more than, at the most, two time horizons.

We therefore introduce the Bandspectrum Cointegration Estimator (BSCE). The difference between the BSCE and the bandspectrum regression is that the BSCE uses non-stationary panel data. The panel is used to pool individuals such that the number of long run observations is increased. The BSCE, like the bandspectrum regression, is estimated in the

---

[5] The VECM model is specified in Tables 4 and 7.

frequency domain since it is relatively easier to identify the various horizons in this domain compared to the time domain. Unlike the bandspectrum regression, however, the BSCE takes the first difference of all time series before they are transformed to the frequency domain. As long as the time series are integrated I(d) where $0.5 \leq d < 1.5$, taking the first difference generates a new stationary time series which can be used without the risk of obtaining spurious results.

The BSCE is based on the following estimation procedure;

i.   Test the integration order of the time series. If the time series are stationary Engle's bandspectrum regression can be used.

ii.   If the time series are non-stationary, take first difference and transform the first differenced data to the frequency domain.

iii.   Estimate the model for each individual frequency or frequency band, and test whether one parameter vector applies to many frequencies.

iv.   If the same parameter applies to many frequencies, combine these into one frequency component and redo the regressions. The parameter estimates are more efficient if they are estimated using a frequency band that is as wide as possible.

4. Fourier and Wavelet Analysis

It is easier to estimate models with many horizons in the frequency domain than in the time domain, because each horizon is represented by its own frequency or set of frequencies. To transform a time series to the frequency domain the Fourier transform is often used, but this is not the only transform that can be employed. An alternative to the Fourier transform is the wavelet transform. The difference between the Fourier transform and the wavelet transform is that the latter combines time and frequency resolution while the Fourier transform only contains frequency resolution[6]. The combination of time and frequency resolution is important when the time series contains non-recurring events such as outliers or structural breaks. The lack of time resolution can otherwise cause the transform to misrepresent the time series in the frequency domain.

---

[6] The wavelet transform transforms a time series to the wavelet domain and not to the frequency domain. The difference is that the wavelet domain combines time and frequency resolution. Since the wavelet transform coefficients contain frequency resolution this presentation will refrain from calling it the wavelet domain and only refer to the frequency domain for simplicity. The reader should, however, be aware of the differences, see Percival and Walden (2006).

It is not uncommon for the economy to go through different regimes over time, for example due to changes in institutions or policy. The combination of time and frequency resolution is therefore important. Time resolution can be introduced into the Fourier transform by using a windowed Fourier transform, which is based on transforming a smaller sub-sample instead of using the entire sample. The transform is obtained by setting a window in the time domain and letting it slide along the time axis. The time series is transformed one window at a time instead of transforming the entire time series all at once. The time series only has to be stationary within each window, but can be non-stationary between windows. This method is sensitive to the size of the window, and many different window sizes have to be used to test the robustness of the transform. This increases the complexity of the analysis.

The wavelet basis functions form a complete orthonormal transform and they are, furthermore, local in time, which means that they start at a given point in time and end at a given point in time. The sine and cosine functions, which the Fourier analysis uses, oscillate along the time axis to infinity and are hence global. Let $\mathcal{W}$ be a (T×T) matrix which contains the wavelet basis functions (for more information about the basis functions see Percival and Walden 2006). The wavelet transform coefficients can be obtained through

$$w = \mathcal{W}x, \tag{20}$$

where x is a (T×1) vector which contains the time series x_t, $t=1,..T$. Since the wavelet functions form an orthonormal basis the inverse wavelet transform is given by the transpose of the transform matrix such that

$$x = \mathcal{W}^T w \tag{21}$$

The transform matrix and the vector with transform coefficients can be partitioned into different scales $j=1,..,J$. This implies that the transform matrix can be decomposed as

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_{J-1} \\ \mathcal{W}_J \\ \mathcal{V}_J \end{bmatrix}, \tag{22}$$

where $\mathcal{W}_j$ is a (T/2^j×T) matrix and $\mathcal{V}_j$ is (1×T). The vector with transform coefficients can accordingly be decomposed as

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{J-1} \\ w_J \\ v_J \end{bmatrix},$$

(23)

where w_j is a vector with $T/2^j$ transform coefficients. The individual scales are interesting because the wavelet basis functions can be interpreted as representing a band pass filters. The transform coefficients at scale j can therefore be interpreted as representing the time series x at frequency[7]

$$\left[\frac{1}{2^{j+1}}, \frac{1}{2^j} \right].$$

(24)

The transform matrix $\mathcal{V}_J$ can be interpreted as a low pass filter while the transform matrixes, $\mathcal{W}_j$, are high pass filters. v_J, which is called the scaling coefficient, thus represents the time series at frequency

$$\left[\frac{1}{0}, \frac{1}{2^{j+1}} \right].$$

(25)

Besides representing a particular frequency the individual transform coefficients within a given scale also represents parts of the time series at given points in time. The individual transform coefficients within a scale is the difference between two (weighted) averages of the time series x. The wavelet transform coefficients thus both contains both time and frequency resolution. However, it is not possible to obtain time resolution and maintain the same detailed frequency resolution that the Fourier transform contains. The wavelet transform thus contains less detailed frequency information than the Fourier transform, but in addition it also contains time resolution, which the Fourier transform does not.

The combination of time and frequency resolution, and the properties of the wavelet basis functions, makes the wavelet transform ideal to study non-stationary time series. It can be shown that the wavelet filter can de-correlate non-stationary time series (see Tewkin and Kim 1992, and Craigmile *et.al* 2005). This implies that we can decompose non-stationary time series and use the transform coefficients to represent the time series at different frequencies. The ODFT cannot decompose a non-stationary time series into frequencies components and requires that the time series are stationary (see Brockwell and Davis 1991 and Phillips 1999).

---

[7] We only need to consider a unit interval of frequencies, and for a real valued time series we only need to consider half of that interval. For more information see Brockwell and Davis (1991)

In this paper we use the DWT when we simulate the non-stationary time series in Section 5 and both the DWT and the ODFT when we estimate the BSCE. BSCE uses stationary first differenced data and the ODFT can thus be applied to transform the time series data to the frequency domain, but it can not be used in the first step of the simulations.

For more information about the DWT see, for example, Percival and Walden (2006) and Crowley (2007) and for more information about the frequency domain see Brockwell and Davis (1991).

5. Simulations

In this section the small sample properties of the BSCE are evaluated and compared to common time domain estimators – the OLS levels estimator, OLS first difference estimator and the VECM. It is also compared to the frequency domain estimator NBFDLS, which uses levels data.

Two sets of simulations are carried out; one where there are three horizons and one where there is just one horizon. The number of long run observations is limited in a small sample from just one individual, and we therefore use a panel data model where we pool five individuals[8].

5.1 Simulations with Three Horizons

The first set of simulations uses a bi-variate fixed effects panel data model. Following, in spirit, Englund *et al* (1992) and Friedman and Kuznets (1954) the data generating process contains three components, a short run shock component, a medium run business cycle component and a long run cointegration relationship. To simulate the time series the following procedure is used;

    i.   Letting i denote the individual and t denote time, we simulated x_{it} as a random walk in the time domain

$$x_{it} = x_{it-1} + \eta_{it} \qquad\qquad (26)$$

        where the shocks, η_{it}, are independent and distributed $N(0,1)$. Since x_{it} follows a random walk, it is a non-stationary I(1) process.

---

[8] We also considered using just one individual, but all estimators had problems estimating the cointegration vector in the case when the data generating process contained three time horizons.

ii.   The second step is to transform x_{it} to the frequency domain[9]. This is achieved by using the discrete wavelet transform and the Daubechie (6) wavelet. Due to the de-correlating property of the Daubechie wavelets for non-stationary time series, it is possible to decompose the time series on a frequency basis without miss-representation.

iii.   Once x_{it} has been transformed we generate y_{if} in the frequency domain, where f denotes frequency, using the following data generating process[10];

$$y_{if} = \begin{cases} \varepsilon_{if} & |f| \in \left[\dfrac{1}{8},\dfrac{1}{2}\right] \\ 0.5x_{if} + \varepsilon_{if} & |f| \in \left[\dfrac{1}{64},\dfrac{1}{8}\right) \\ x_{if} + v_{if} & |f| \in \left[0,\dfrac{1}{64}\right) \end{cases} \tag{27}$$

where

$$v_{if} = \varepsilon_{if} + \lambda_i \tag{28}$$

The residuals ε_{it} are drawn from a normal distribution and are independently distributed $N(0,\sigma_i^2)$ with $\sigma_i^2 \sim U(0.5,1.5)$. The individual fixed effects follow a uniform distribution; $\lambda_i \sim U(0,4)$. Each individual is independent of the others.

iv.   Once y_{if} has been generated it is transformed to the time domain using the inverse discrete wavelet transform.

v.   The OLS levels, OLS first difference and the VECM models are now estimated in the time domain. The non-stationary levels data is transformed to the frequency domain so we can estimate the NBFDLS. Finally we take the first differences of all x and y in the time domain to enable us to use the BSCE.

The data generating process we used in the simulations has the following interpretation; in the short run, there is no correlation between x_{it} and y_{it} because they are both affected by uncorrelated shocks. In the medium run (the business cycle) there is a relationship but it is weaker than the long run relationship. If the model was specified on quarterly data, the first 1-

---

[9] We are actually transforming into the wavelet domain, which contains both frequency and time resolution. Because we use the frequency resolution of the transform coefficients in the simulations we call this the frequency domain for the sake of simplicity.

[10] The negative frequencies have the same data generating process as the positive frequencies since x_{it} is a real valued time series.

4 quarters are the short run, the following 4-32 quarters represents the business cycle, and the remaining 32 quarters and beyond is the long run.

The number of observations for each individual was first set to T=256[11] and then to T=512. For one individual there are eight long run observations when T=256 and sixteen when T=512. We use five individuals in the panel – we also estimated panels with fewer individuals and found that increasing the number of individuals from one to three improved the results for all estimators, but there were no significant gains for any estimator when the panel contained more than three indididuals, but to be on the conservative side we use five individuals in the simulations. In total there are therefore 40 long run observations when T=256 and 80 when T=512.

The number of observations per individual (T=256 and T=512) was determined by the fact that the discrete wavelet transform which requires 2^J observations to transform the data, where J is an integer. If the number of observations in an empirical analysis does not equal to 2^J, it is always possible to pad the data by adding observations (for example the mean) until the restriction is met. Because the discrete wavelet transform contains time resolution it is possible to remove those transform coefficients that are affected by the data padding once the time series have been transformed. The use of data padding therefore has no effect on the regression results, and was therefore ignored in this study.[12, 13]

5.2 Simulation Results

5.2.1 Simulation Results – Three Horizons

Each set of simulations (T=256, T=512) was repeated 5000 times. The average parameter estimates and their empirical standard deviations are presented in Table1 for the BSCE. Table 2 shows the bandspectrum regression that has been applied to the non-stationary levels data,

---

[11] To simulate the time series we set T=4096 and then use 256 and 512 of these observations. This is done to avoid problems with boundary coefficients.

[12] When the time series were simulated we used the Daubechie (6) wavelet due to its de-correlating property for non-stationary time series. Because the BSCE is estimated on first differenced data we do not need this property when we employ the BSCE. We therefore used the Daubechie (4) for this estimator because the Daubechie (6) wavelet generates more boundary coefficients than the Daubechie (4) wavelet.

[13] In these simulations the DWT-BSCE is estimated with a Daubechie (4) wavelet. When his is used some wavelet transform coefficients are affected by boundary conditions. These are removed from the estimation (see Percival and Walden 2006) so as not affect the estimation results. It implies that the BSCE is estimated on 6 long run observations when T=256 and 14 when T=512 for each individual. The BSCE is therefore estimated on fewer observations than the OLS levels, OLS first difference, VECM and Frequency Domain Least Squares estimators.

the parameter estimate for the frequency band 0 to 1/64 in this table represents the NBFDLS. Table 3 presents the results for the OLS levels and OLS first difference estimators. The VECM results are given in Table 4.

Both the DWT-BSCE[14] and the ODFT-BSCE find that there are three different parameter vectors when using the testing procedure described at the end of Section 3. The DWT-BSCE parameter estimates are less biased, however, and more efficient than the ODFT-BSCE estimates. For example, the true cointegration parameter (long run) is 1. The DWT-BSCE estimates are 0.989 (T=256) and 0.992 (T=512) while the ODFT-BSCE estimates are 0.955 and 0.945 respectively. The standard deviations are 0.018 and 0.007 for the DWT-BSCE and 0.060 and 0.032 for the ODFT-BSCE.

The differences between the DWT-BSCE and the ODFT-BSCE are even greater for the medium run results. The DWT-BSCE parameter estimates are 0.501 and 0.503 depending on the sample size, compared to the true value of 0.5. The corresponding parameter estimates from the ODFT-BSCE are 0.429 and 0.431, these estimates are also less efficient than the DWT-BSCE estimates. There is no appreciable difference for the short run results, however.

In Table 2, the DWT and the ODFT are applied to the levels data instead of the first differenced data. The long run results from the ODFT in Table 2 represent NBFDLS. The loss of time resolution is noticeable for the ODFT, which therefore contains relatively larger biases than the DWT for the cointegration parameter estimate. The ODFT is furthermore, not capable of distinguishing between the other horizons (short-run and medium-run). The DWT identifies all three horizons, but the parameter estimates for the medium run is biased.

An interesting result of the simulations is that the OLS levels estimator, in Table 3, has a relatively large bias. The estimated common parameter vector should converge in probability to the cointegration parameter due to the non-stationarity of the data, but there are too few long run observations for this to happen. The cointegration parameter estimate has a bias of -0.061 and -0.031 respectively compared to the true value, which is greater than the bias for the DWT-BSCE. The parameter estimates are furthermore less efficient than the DWT-BSCE estimates despite the fact that the OLS levels estimates use non-stationary levels data, while the BSCE is estimated using first differenced data.

The VECM estimates of the cointegration vector in Table 4 are slightly less biased and more efficient than the OLS levels estimates, but inferior to the DWT-BSCE. The bias of the

---

[14] DWT-BSCE denotes the bandspectrum cointegration estimator using a DWT, and ODFT-BSCE the estimator using the ODFT.

VECM cointegration parameter estimate is -0.033 and -0.015 and the respective standard deviations are 0.021 and 0.009. The other parameters from the VECM regression represent the shorter horizons, which, however, are difficult to interpret in relation to the true data generating process. If the data generating process had contained just two horizons these parameters would have represented the short run, but now they represent a mixture of the short run and the medium run.

The parameter estimates from the OLS first difference estimator are 0.136 and 0.138. The theoretical parameter value for these estimates can be obtained from equation (16), and is 0.141. The OLS estimator, converges, as expected, to the theoretical value, but contains no relevant information about either horizon.

Introducing more than one horizon into the data generating process obviously creates problems for the usual time domain estimators, as is evident from the results. The parameter estimates are biased and less efficient than for the DWT-BSCE. Figures 1-8 plots the empirical distribution of the respective parameter estimates. Figures 1 and 2; Panel A shows the distribution of the DWT/ODFT-BSCE cointegration vector estimates and Panel B the medium run parameter estimates for $T=256$, and Figures 5 and 6 for $T=512$. These estimates appear to have a symmetric distribution. Similar figures for the OLS levels and VECM estimators are given in Figures 3 and 4 for $T=256$ and 7 and 8 for $T=512$. As is apparent in the figures, the distributions of these estimates are skewed to the left. An estimate of the skewness is given in Table 8; the skewness of the OLS levels estimator is about -1.25 and for the VECM cointegration parameter estimates about -1.5. This skewness is a result of two effects; the small sample, and the fact that the medium run and short run parameters are smaller than the long run parameter. There is also a slight skewness for the cointegration vector for the DWT-BSCE (about -0.25) and ODFT-BSCE (about -0.35) but no skewness for the medium and short run parameter estimates.

5.2.2 Simulation Results - Common Parameter Vector

To see how the BSCE works when there is only one horizon we can let the true data generating process be given by equation (3). Table 5 presents the results when there is one common parameter vector for all frequencies for the BSCE, Table 6 for the OLS estimators, and Table 7 the VECM estimators.

All the estimators are unbiased, but the OLS levels estimator is the most efficient; it has a standard deviation of 0.006 ($T=256$) and 0.003 ($T=512$). The DWT-BSCE is slightly less efficient 0.015 and 0.007, but the differences are very small. The OLS first difference

estimator is also unbiased but inefficient, the standard deviations are 0.037 and 0.028 respectively.

It is interesting to note, however, that the loss off efficiency is relatively small for the OLS first difference estimator compared to the OLS levels estimator. It is also interesting that the adjustment parameter in the VECM model, α_1 is equal to -1. This implies that there are no persistent deviations from the equilibrium cointegration relationship. The adjustment parameter was much smaller when the data generating process also contained a medium and a short run vector, α_1=-0.257 (T=256). This result indicates that we may infer that the true data generating process contains more than one horizon when the adjustment parameter is smaller than 1 in absolute value. The same conclusion may be drawn when the OLS first difference estimator and the OLS levels estimator yield significantly different parameter estimates. When either of the above two results is obtained in an empirical analysis, the BSCE may be the right estimator to consider.

6. Conclusions

Economic theory commonly distinguishes between different time horizons, each with its own dynamics and relationships. Engle (1974) proposed a bandspectrum regression for such models when the data is stationary. This paper considers a Bandspectrum Cointegration Estimator for models with non-stationary panel data. The BSCE estimates the model in the frequency domain using first differenced data to avoid problems with spurious results.

We evaluate the small sample properties of the BSCE in a simulation study by comparing it to common time domain and frequency domain estimators – OLS levels estimator, OLS first difference, VECM and NBFDLS. The results of the simulations show that the BSCE successfully identifies all the horizons in the data and estimates a separate parameter vector for each. We also find that the BSCE estimates of the cointegration vector are less biased and more efficient than the estimates from any of the other estimators when the data generating process contains three time horizons. All estimators are unbiased when there is only one time horizon, and the BSCE is, as expected, slightly less efficient than estimators that use non-stationary levels data. This loss of efficiency is very small, however.

References

Brockwell P.J. and David R.A., 1991. Time Series: Theory and Method. Springer-Verlag, New York.

Craigmile P.F., Guttorp P., and Percival D.B., 2005. Wavelet-Based Parameter Estimation for Polynomial Contaminated Fractionally Differenced Processes. IEEE Transactions on Signal Processing 53 (8), 3151-3161.

Crowley P.M., 2007. A Guide to Wavelets for Economists. Journal of Economic Surveys 21 (2), 207-415.

Davidson R. and MacKinnon J.G., 2004. Econometric Theory and Methods. Oxford University Press, New York, USA.

Engle R.F., 1974. Bandspectrum Regression. International Economic Review 15(1), 1-11.

Engle R.F. and Granger C.W.J., 1987. Co-Integration and Error Correction: Representation, Estimation and Testing. Econometrica 55(2), 251-276.

Englund P., Persson T. and Svensson L.E.O., 1992. Swedish Business Cycles: 1861-1988. Journal of Monetary Economics 30, 343-371.

Friedman M. and Kuznets S., 1954. Income from Independent Professional Practice. National Bureau of Economic Research, New York USA.

Granger C.W.J., 1966. The Typical Spectral Shape of an Economic Variable. Econometrica 34, 150-161.

Granger C.W.J., 1980. Long Memory Relationships and the Aggregation of Dynamic Models. Journal of Econometrics 14, 227-238.

Granger C.W.J and Joyeux R., 1980. An Introduction To Long-Memory Time Series Models and Fractional Differencing. Journal of Time Series Analysis 1.

Granger C.W.J. and Newbold P., 1974. Spurious Regressions in Econometrics. Journal of Econometrics 2, 111-120.

Johansen S., 1988. Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control 12, 231-254.

Johansen S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica 59, 1551-1580.

King R.G. and Watson M.W., 1996. Money, Prices, Interest Rates and the Business Cycle. The Review of Economics and Statistics 78(1), 35-53.

Percival D.B. and Walden A.T., 2006. Wavelet Methods for Time Series Analysis. Cambridge University Press, New York USA.

Phillips P.C.B, 1991. Spectral regression for cointegrated time series. Cowles Foundation Paper 796.

Phillips P.C.B., 1999. Discrete Fourier transforms of Fractional Processes. Mimeo Yale University.

Ramsey J.B., 2002. Wavelets in Economics and Finance: Past and Future. C.V. Starr Center for Applied Economics 2002-02.

Ramsey J.B. and Lampart C., 1998. Decomposition of Economic Relationships By Timescale Using Wavelets. Macroeconomic Dynamics 2, 49-71.

Robinson P.M. and Marinucci D., 1998. Semiparametric frequency domain analysis of fractional cointegration. LSE Discussion Paper No. EM/98/348.

Tewfik A. and Kim M., 1992). Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion. IEEE Transactions on Information Theory 38, 904-909.

Yule G.U., 1926. Why do We Sometimes Get Nonsense-Correlations between Time-Series? – A Study in Smaplinand the Nature of Time-Series. Journal of Royal Statistical Society, 89(1), 1-63.

Table 1: Three Horizons – BSCE

Data Generating Process		DWT		ODFT	
Frequency band	*True β*	*Parameter Estimate*	*Std. Dev.*	*Parameter Estimate*	*Std. Dev.*
		T=256			
0	1	0.989	0.018	0.955	0.060
1/64	0.5	0.501	0.030	0.429	0.047
1/8	0	0.002	0.114	0.030	0.142
		T=512			
0	1	0.992	0.007	0.945	0.032
1/64	0.5	0.503	0.020	0.431	0.032
1/8	0	0.000	0.081	0.023	0.105

Table 2: Three Horizons – Non-Stationary Levels Data

Data Generating Process		DWT		ODFT[1]	
Frequency band	True β	Parameter Estimate	Std. Dev.	Parameter Estimate	Std. Dev.
		T=256			
0	1	0.998	0.007	0.975	0.025
1/64	0.5	0.549	0.031	0.611	0.069
1/8	0	0.022	0.096	0.462	0.205
		T=512			
0	1	0.999	0.002	0.987	0.010
1/64	0.5	0.550	0.019	0.581	0.049
1/8	0	0.000	0.053	0.451	0.191

Note:

1. The low frequency estimator (frequency 0 to 1/64) is the Narrow Band Frequency Domain Least Squares Estimator.

Table 3: Three Horizons – OLS Estimators

Estimator	True β	T=256 Parameter Estimate	T=256 Std. Dev.	T=512 Parameter Estimate	T=512 Std. Dev.
OLS Levels	1	0.939	0.032	0.969	0.016
OLS First Difference	0.141	0.136	0.220	0.138	0.108

Table 4: Three Horizons – VECM[1]

$$\Delta y_t = \alpha_1 (y_{t-1} - \beta x_{t-1}) + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \gamma_3 \Delta x_{t-1} + \gamma_4 \Delta x_{t-2}$$

$$\Delta x_t = \alpha_2 (y_{t-1} - \beta x_{t-1}) + \gamma_5 \Delta y_{t-1} + \gamma_6 \Delta y_{t-2} + \gamma_7 \Delta x_{t-1} + \gamma_8 \Delta x_{t-2}$$

	T=256		T=512	
Parameter	*Parameter Estimate*	*Std. dev.*	*Parameter Estimate*	*Std. dev.*
α_1	-0.257	0.059	-0.242	0.039
α_2	0.300	0.043	0.295	0.031
β	0.967	0.021	0.985	0.009
γ_1	-0.565	0.058	-0.578	0.039
γ_2	-0.259	0.038	-0.265	0.026
γ_3	0.124	0.062	0.133	0.042
γ_4	0.115	0.056	0.119	0.039
γ_5	-0.117	0.041	-0.113	0.029
γ_6	-0.012	0.031	-0.010	0.021
γ_7	0.143	0.037	0.142	0.026
γ_8	0.103	0.037	0.104	0.026

Note:

1. The theoretical parameter estimate for the cointegration relationship (β) is 1. The other parameter estimates represent a mixture of the short and the medium run.

Table 5: Common Parameter Vector – BSCE

Data Generating Process		True β	DWT		ODFT	
Frequency band			Parameter Estimate	Std. Dev.	Parameter Estimate	Std. Dev.
T=256						
0	1/64	1	1.000	0.015	1.000	0.020
1/64	1/8	1	1.001	0.041	1.000	0.037
1/8	1/2	1	0.999	0.114	1.000	0.086
T=512						
0	1/64	1	1.000	0.007	1.000	0.020
1/64	1/8	1	1.000	0.028	1.000	0.037
1/8	1/2	1	1.000	0.080	1.000	0.088

Table 6: Common Parameter Vector - OLS Estimator

Estimator	True β	T=256		T=512	
		Parameter Estimate	*Std. Dev.*	*Parameter Estimate*	*Std. Dev.*
OLS Levels	1	1.000	0.006	1.000	0.003
OLS First Difference	1	1.000	0.037	1.000	0.028

Table 7: Common Parameter Vector – VECM

$$\Delta y_t = \alpha_1 (y_{t-1} - \beta x_{t-1})$$
$$\Delta x_t = \alpha_2 (y_{t-1} - \beta x_{t-1})$$

| Parameter | T=256 | | T=512 | |
	Parameter Estimate	Std. dev.	Parameter Estimate	Std. dev.
α_1	-1.004	0.086	-1.003	0.063
α_2	-0.024	0.063	-0.013	0.045
β	0.995	0.011	0.998	0.005

Table 8: Three Horizons - Skewness

	Discrete Wavelet Transform[1]		Discrete Fourier Transform[1]	
Frequency band	$T=256$	$T=512$	$T=256$	$T=512$
BSCE				
0 — 1/64	-0.266	-0.187	-0.355	-0.377
1/64 — 1/8	0.030	0.035	-0.059	-0.016
1/8 — 1/2	0.008	0.010	0.000	-0.012

	$T=256$	$T=512$
OLS		
Levels	-1.255	-1.263
First Difference	-0.015	-0.003
VECM		
Cointegration Vector	-1.506	-1.464

Note:

1. Estimated using first differenced data

Figure 1: DWT - BSCE *T=256, I=5*

Panel A: Cointegration

Panel B: Medium Run
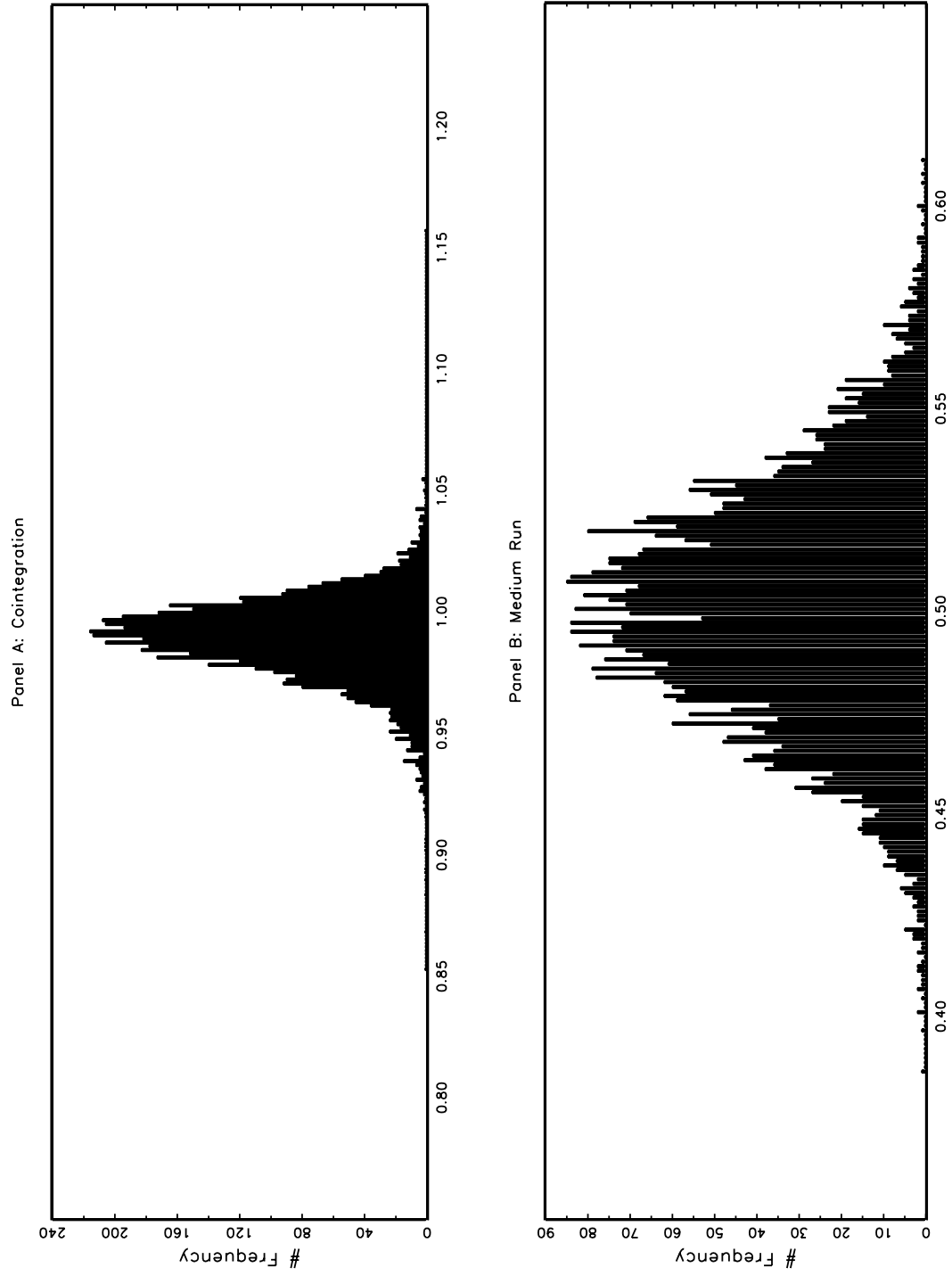
Figure 2: ODFT - BSCE $T=256, I=5$

Panel A: Cointegration

Panel B: Medium Run

Figure 3: OLS *T=256,I=5*

Panel A: OLS Levels

Panel B: OLS First Difference

29
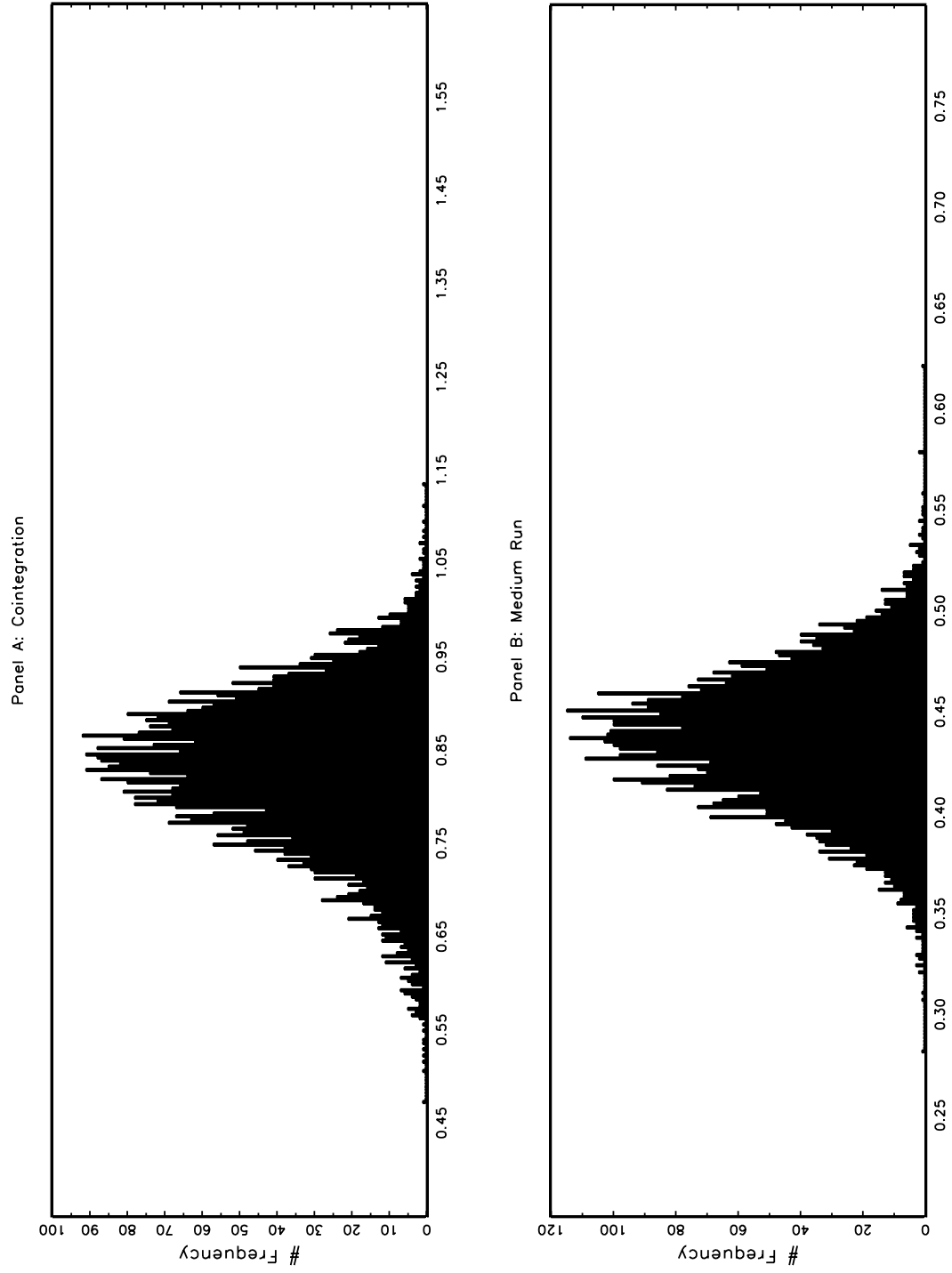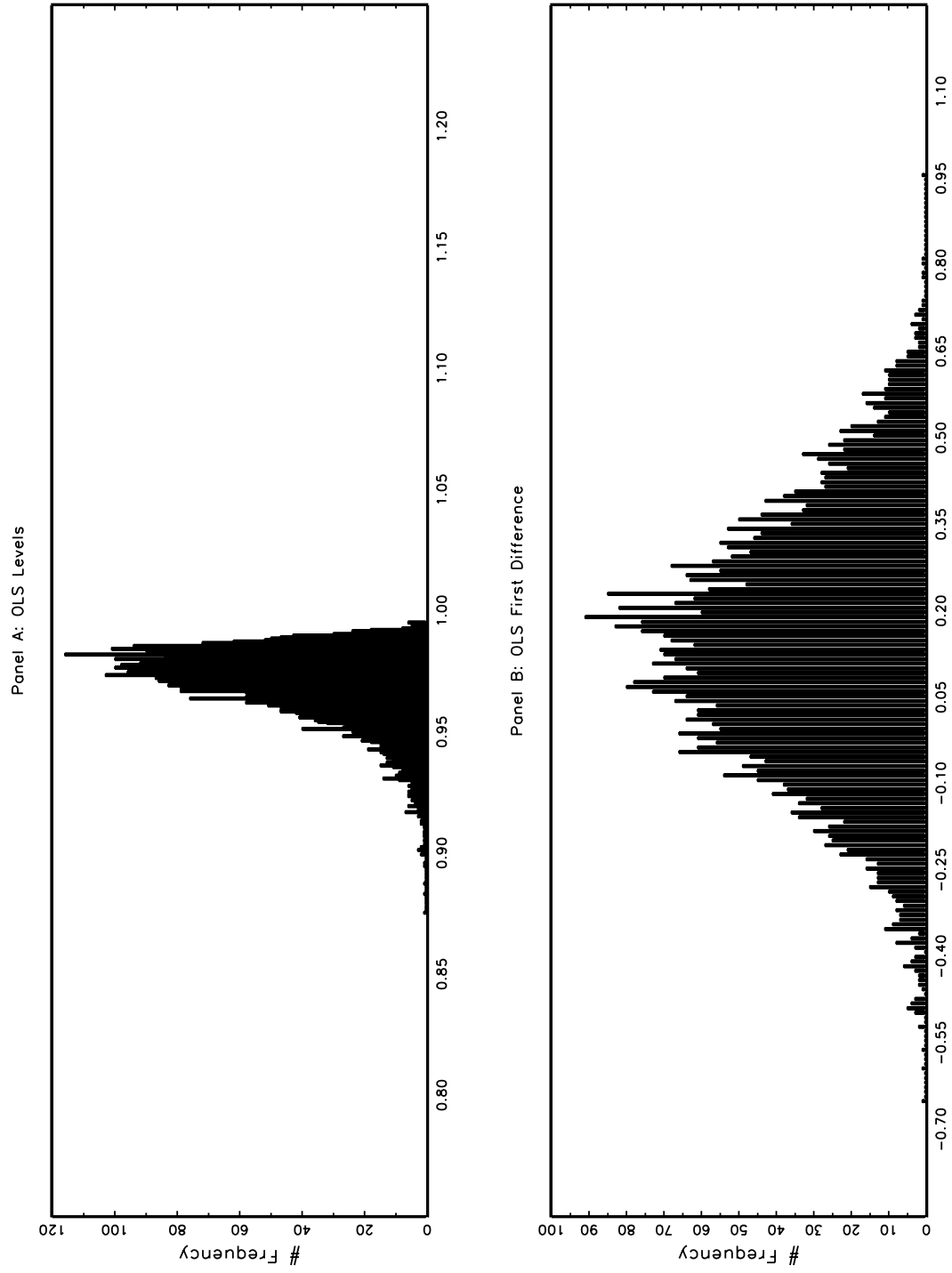
Figure 4: VECM $T=256, I=5$

Figure 5: DWT - BSCE *T=512, I =5*

Panel A: Cointegration

Panel B: Medium Run
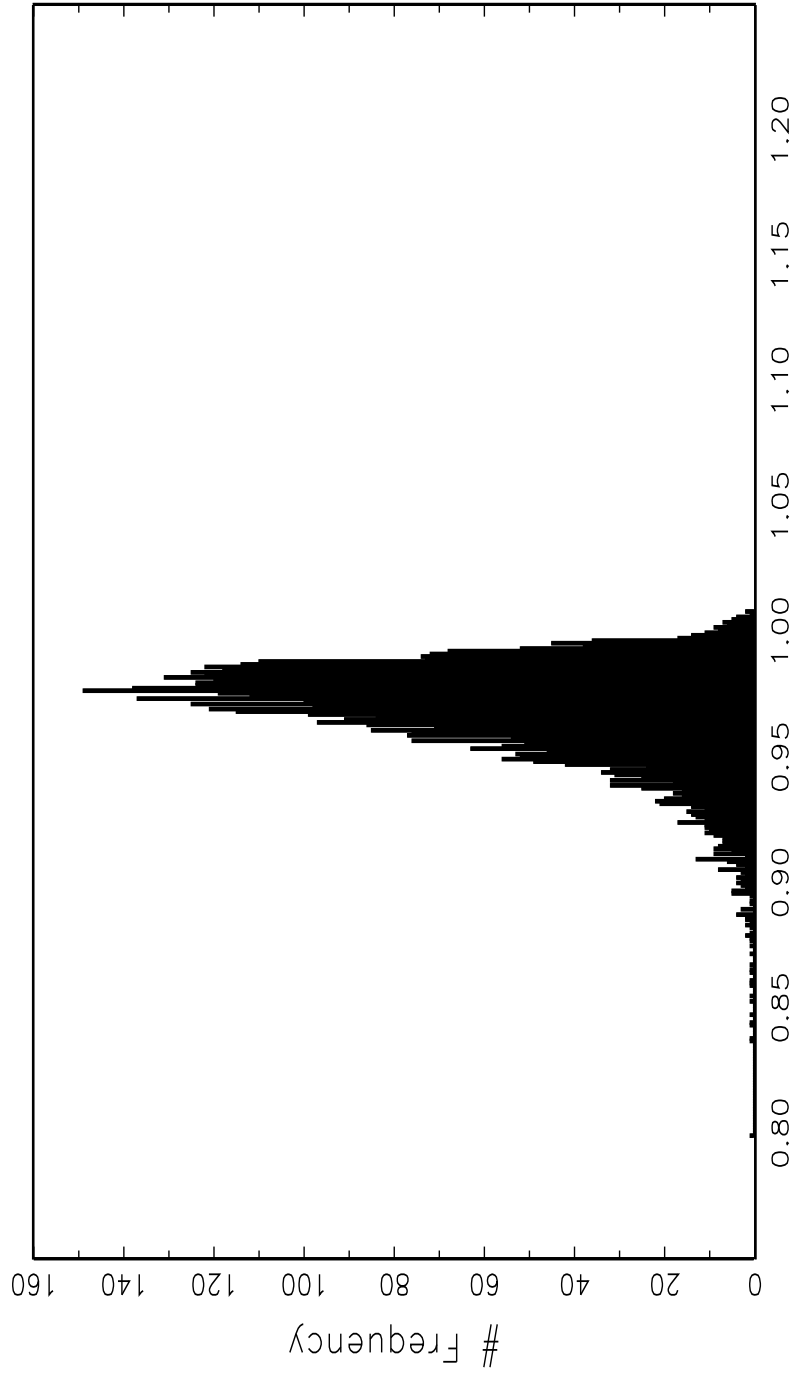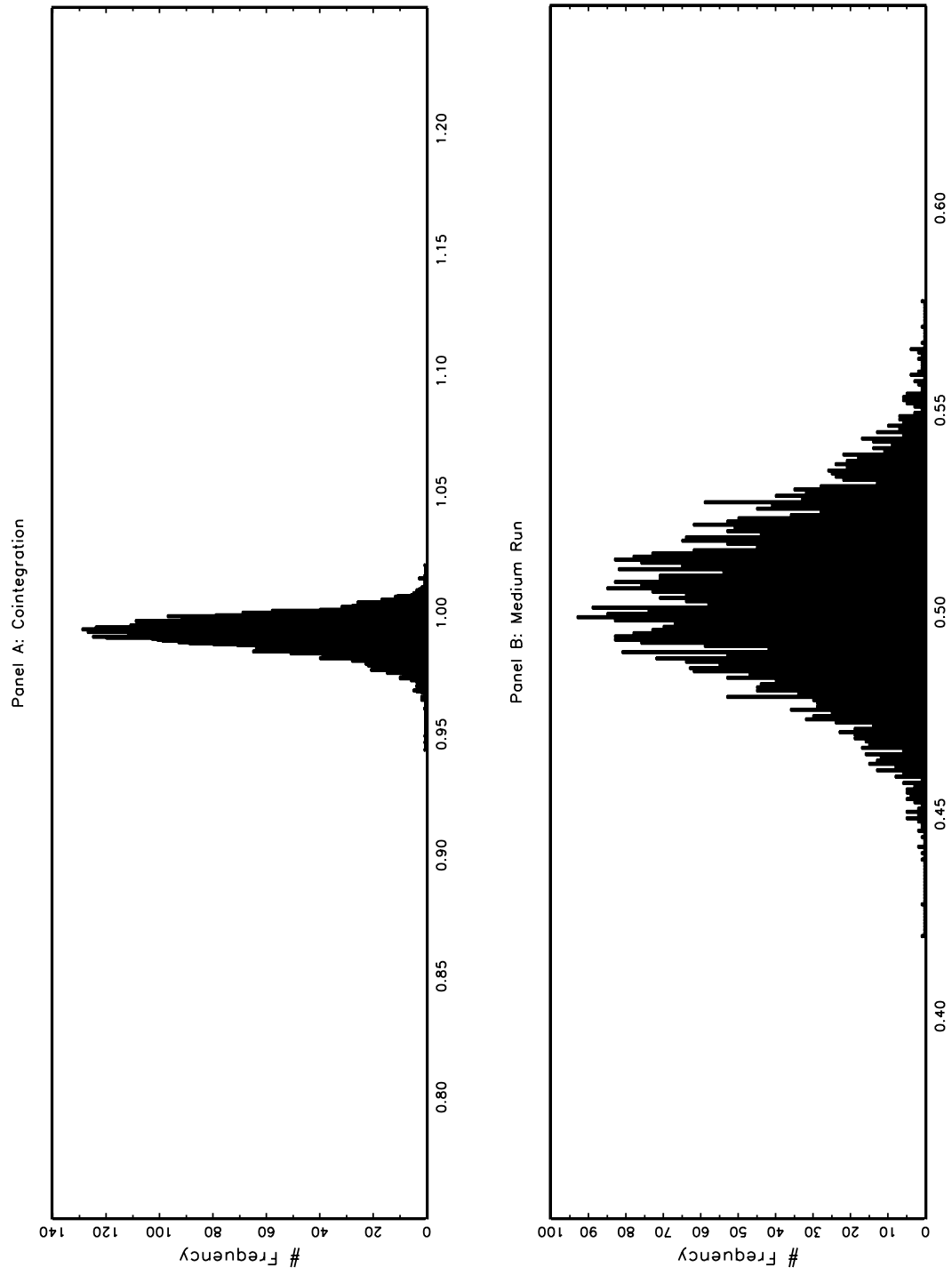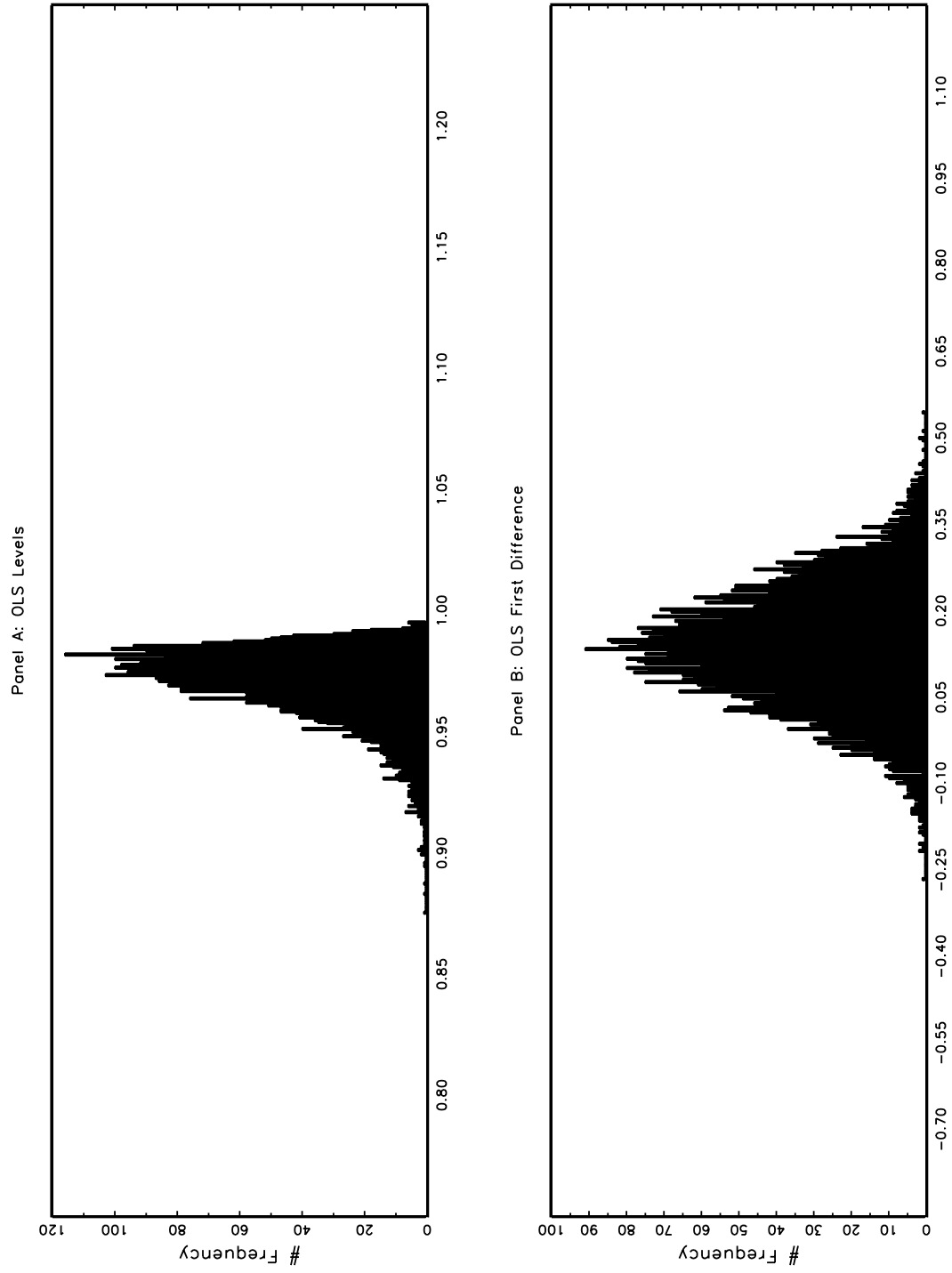
Figure 6: ODFT - BSCE T=512, I=5

Panel A: Cointegration

Panel B: Medium Run

Figure 7: OLS *T=512,I=5*

Panel A: OLS Levels

Panel B: OLS First Difference

Figure 8: VECM *T*=512, *I*=5