Generalized Spectral Estimation

Jeremy Berkowitz*

June 12, 1996

Abstract

This paper provides a framework for estimating parameters in a wide class of dynamic rational expectations models. The framework recognizes that dynamic RE models are often meant to match the data only in limited ways. In particular, interest may focus on a subset of frequencies. Thus, this paper designs a frequency domain version of GMM. The estimator has several advantages over traditional GMM. Aside from allowing band-restricted estimation, it does not require making arbitrary instrument or weighting matrix choices. The general estimation framework also includes least squares, maximum likelihood and band restricted maximum likelihood as special cases.

Key Words: Estimation, Frequency Domain, Misspecification
JEL Classification: C13, C22

* Federal Reserve Board, Mail Stop 61A, Washington, D.C. 20551. Telephone: (202) 736-5581. Email: m1jmb02@FRB.GOV. I would like to thank Frank Diebold, Lee Ohanian, Valentina Corradi and Jin Hahn for helpful suggestions. The view presented are solely those of the author and do not necessarily represent those of the Federal Reserve Board or its staff.

1 Introduction

This paper develops frequency domain techniques for estimating dynamic rational expectations models. This approach allows, models to be estimated and tested over a subset of frequencies, such as business cycle frequencies, seasonal frequencies, or long horizons. The techniques described in this work are also particularly useful in allowing researchers to deal squarely with high frequency measurement error.

It is natural for researchers interested in avoiding high frequency noise or in matching particular cyclical behavior to carry out estimation and evaluation of such models in the frequency domain. The frequency domain provides an orthogonalization of the fluctuations in the observed data. Engle (1974) introduced band spectral regression as a means to assess the relationship between economic variables at specific frequencies. In that work, the criterion being minimized was restricted to linear least squares.

Generalized spectral estimation (GSE) allows for a much wider class of minimization criteria than was previously possible. The GSE framework includes a new class of estimators which I will call 'whitening estimators', as well as least squares, band spectrum regression, and 'Whittle likelihood' estimation (which is asymptotically maximum likelihood).[1]

This paper builds on Diebold, Ohanian and Berkowitz (1995), who propose techniques for estimating and evaluating dynamic rational expectations models in the frequency domain. Their framework allows for parameter estimation and model assessment in a very general setting. Parameters are estimated by minimizing distance between spectra of observed data and model-generated data. Distance may be defined by the user and may focus on any relevant subset of frequencies. A shortcoming of this framework is that in all but the simplest cases, the model must be approximated and simulated in order to carry out the estimation.

Since, in general, analytic solutions are not available for nonlinear dynamic equilibrium models, one must choose an approximate solution method. Furthermore, in order to proceed with estimation, the approximation to the model must be carried out at each parameter configuration. That means the model must be simulated hundreds if not thousands of times. Solution methods which can be chosen arbitrarily 'close' to the true model (such as discretizing the parameter space) are generally precluded because of the extreme computational intensity associated with simulating the model only once. For estimation, then, faster and less accurate solution methods are required. To the extent that the approximate solution differs from the true model, the Diebold, Ohanian and Berkowitz (1995) estimated parameters will differ from the parameters which minimize loss. It is difficult to make general statements regarding

[1]For a discussion of band-restricted maximization of the Whittle likelihood function see Engle (1980), Diebold, Ohanian, and Berkowitz (1995).

this sort of approximation error. However, Taylor and Uhlig (1990), in a comparison of 14 approximation methods applied to the stochastic growth model, concluded that "the simulated sample paths generated by the different solution methods have significantly different properties."

The generalized spectrum estimator will allow for model estimation, inference and evaluation without requiring an approximate solution of the model. We accomplish this by imposing moment conditions given by the model and then minimizing deviations from these conditions in the frequency domain. It is thus very much in the spirit of generalized method of moment and other minimum distance estimators. For whitening estimators, we impose moment conditions on the residuals which require that the residuals are 'close' to white noise.

Section 2 defines the generalized spectral estimator and presents some special cases for illustration. Section 3 illustrates GSE estimation by presenting the results of a Monte Carlo experiment. In the experiment, I maintain the realistic assumption that the true model is unknown. Section 4 concludes.

2 Generalized Spectral Estimation

The Euler equation implied by a typical rational expectations model can be written as

$$E(g(y_t, \theta_0) \mid \Omega_t) = \mathbf{0}, \tag{1}$$

where g$(\cdot,\cdot)$ is a function given by model's first order conditions, y_t is an Tx1 vector of observable data, θ_0 is a vector of parameter values, and Ω_t is the ($\sigma-$algebra defined by) agent's time t information set. $g(y_t, \theta)$ is sometimes called the Euler residual. Equation 1 says that the Euler residual has a zero conditional mean. It implies that for any rx1 instrument x_{t-1}, in the agent's time t information set,

$$E\left(g(y_t, \theta_0) \otimes \begin{bmatrix} 1 \\ x_{t-1} \end{bmatrix}\right) = \mathbf{0}. \tag{2}$$

Equation 2 is the familiar basis for GMM estimation. The notion here is that economic agents should be unable to forecast the model residual. Each element of x_{t-1}, each instrument, gives us an additional residual whose mean should be zero.

x_{t-1} is generally taken to be some number of lags of the endogenous variables, y_t. However, consideration of lags of $g(y_t, \theta_0)$ as instruments will lead to some very interesting results. Note that it is *always* true that all lags of $g(y_t, \theta_0)$ are in agents' time t information set. The motivation for doing this is that equation 1 says that the Euler residuals form a martingale difference sequence. Since martingale difference sequences are also white noise, equation 2 says that the Euler residual $g(y_t, \theta_0)$ is multivariate white noise[2].

This fact provides the motivation for a class of GSE estimators, which I will call 'whitening estimators.' The reasoning is as follows. The model residual by design should contain only the unexplainable part of the model in question. So, if we wish to estimate θ_0, we would like to eliminate as much of the predictable dynamics in $g(y_t, \theta_0)$ as possible. These dynamics should be incorporated into the model. Another way of saying this, is we would like to estimate the parameter configuration in such a way as to make the residuals as 'close' as possible to white noise. We are, in effect, trying to whiten the $g(y_t, \theta_0)$. Some examples of whitening estimators will be provided below.

Rewriting equation 3, we have that

$$E\left(g(y_t, \theta_0) \cdot g(y_{t-\tau}, \theta_0)\right) = \mathbf{0},\ \tau = 1, 2, ... \tag{3}$$

For notational simplicity, let $u_t = g(y_t, \theta_0)$, so that u_{jt} is a single Euler residual. Then equation 3 is equivalent to the statement,

$$f_{u_{jt}}(\omega) = k,\ \omega_l = \frac{2\pi l}{T}, l = 0, ... \frac{T}{2}.$$

The spectral density of each u_{jt}, is constant over the entire support, $[0, \pi]$. Then given a finite realization of the residuals, $\{u_{jt}\}_1^T$, we can calculate a consistent estimate of the spectrum, such that $\hat{f}_{u_j}(\omega) \overset{p}{\to} k, \forall\omega$.

In reality, though, $u_t = g(y_t, \theta_0)$, and we do not know θ_0. So we never observe a realization of the true Euler residuals$\{u_{jt}\}_1^T$, but rather at best we can observe $\{u_{jt}(\theta)\}_1^T$, where θ is chosen somehow by the user.

2.1 A Simple Example

In order to fix ideas, consider a simple linear example, an AR(1),

$$y_t = \theta_0 y_{t-1} + u_t.$$

The y_t are observed and assume $u_t \sim (0, 1)$. In GSE notation, $u_t(\theta_0) = g(y_t, \theta_0) = (1-\theta_0 L)y_t$, where L is the lag operator. If we do not know θ_0, we may look at the spectrum of the residual implied by any arbitrary θ, $f_{u(\theta)}(\omega)$, to see whether it is constant across all frequencies. For this arbitrary θ, we can generate a length T residual,

$$u_t(\theta) = (1 - \theta L)y_t$$

[2]By this, I mean that, if $g(y_t, \theta_0)$ is a vector sequence, each element of $g(y_t, \theta_0)$ is white noise.

$$= (1 - (\theta_0 + \alpha)L)y_t$$

$$= u_t - \alpha y_{t-1} = u_t - \alpha \frac{1}{1 - \theta_0 L} u_{t-1} = A(L)u_t.$$

And so the spectrum of this residual $u_t(\theta)$ is,

$$f_{u(\theta)}(\omega) = \left| 1 - \left(\frac{\alpha}{1 - \theta_0 e^{i\omega}} \right) e^{i\omega} \right|^2$$

Then $f_{u(\theta)}(\omega) \neq k$, across frequencies, unless $\alpha = 0$ (that is, at $\theta = \theta_0$).

It is in this sense natural to estimate θ_0, by setting the spectrum $f_{u(\theta)}(\omega)$ 'close' to the spectrum of white noise. In other words, find

$$\hat{\theta}_{GSE} = \arg\min_{\theta} \sum_{\omega_T} C\left(f_{u(\theta)}(\omega)\right),$$

where the loss function C($\cdot$) is a measure of divergence of the spectrum from a flat line.

This approach has some significant advantage relative to Maximum Likelihood and GMM estimation. To continue with the example, ML estimation would proceed by minimizing the distance between the model data, $\{y(\theta)\}_1^T$, and the observed data, $\{y\}_1^T$.

The problem is that in the context of general dynamic equilibrium models the data generating process (the policy function) is not known analytically. The innovation process, $\{u_t(\theta)\}_1^T$, must first be simulated and then the rest of the model data must be generated via an approximate policy function. MLE thus requires an additional layer of approximation.

The advantage of GMM is that one never needs to construct an approximate policy function. Parameters can be estimated using only the observed data. The problem with GMM is that there is no such thing as *the* GMM estimator. The researcher is required to make a choice regarding both instruments and the weighting matrix.

GSE, like GMM, can proceed without requiring an approximate solution of the model. But unlike GMM, the instruments and weighting are not arbitrarily chosen. Rather, the user chooses a frequency band of interest. For example, if the model is designed to explain business cycle fluctuations, the user may specify cycles between 2 to 8 years.

Definition: The generalized spectral estimator is defined as:

$$\hat{\theta}_{GSE} = \arg\min_{\theta} \sum_{\omega_T \in \varpi} C\left(\hat{f}_{u(\theta)}(\omega)\right),$$

where C(·) may include a wide variety of appropriate loss functions, $\hat{f}_{u(\theta)}(\omega)$ is a consistent estimate of the sample spectral density matrix (or possibly the periodogram matrix) of $u_t(\theta) = g(y_t, \theta)$. The summation is taken over the frequencies of interest, represented by the set ϖ. We may appropriately think of the $\hat{f}_{u(\theta)}(\omega_l)$ as a triangular array with $\omega_l = \frac{2\pi l}{T}$, $l = 0, ...\frac{T}{2}$. The maximum frequencies per sample size are $\frac{T}{2}$. This loss function allows for focusing on a subset of frequencies which are judged to be economically relevant. This is not possible in the time domain. To get a better feel for this estimation strategy, we now describe some examples of the loss function C(·). For what follows, we consider only a univariate $u_t(\theta)$.

2.2 Whitening Estimators

We might specify,

$$\hat{\theta} = \arg\min_{\theta} \sum_{\omega} \left(\pi \hat{F}_{u(\theta)}(\omega) - \omega\right)^2, \tag{4}$$

where $\hat{F}_{u(\theta)}(\omega)$ is the spectral distribution or 'cumulative spectrum' of u_t. We have, in effect, inverted the Cramer-Von Mises statistic to obtain an estimator[3]. A few comments are in order here.

1) Rational expectations implies that u_t is white noise, so our moment conditions are that $E\left(\pi \cdot F_{u(\theta)}(\omega)\right) = \omega$. In words, we have that, in population, the spectral distribution function of white noise is the 45° line. This estimator minimizes the L^2 distance between the sample spectral distribution and the population spectral distribution of white noise.

2) This estimator is very much like GMM in that it minimizes the squared deviations from moment conditions.

3) There are an infinite number of valid functions, $C(\cdot)$, with unique minima where u_t is white noise. Durlauf (1991), for example, considers a number of specific distance functions that measure deviations of the cumulative periodogram in the context of testing for white noise.

4) We suppress dependence of the set of ω on the sample size for the remainder of the paper for notational simplicity.

A particularly interesting example of a whitening estimator, which I will call "Spectral-GMM", arises when we take $C(\cdot)$ as follows,

[3] The Cramer-Von Mises statistic is actually $\int_0^1 (BB(z))^2 \, dz$, where BB(z) is a Brownian Bridge.

$$\overset{\wedge}{\theta} = \arg\min_{\theta} \sum_{\omega \in \varpi} \left(\hat{f}_{u1(\theta)}(\omega) - \bar{f}_{u(\theta)} \right)^2, \tag{5}$$

where $\bar{f}_{u(\theta)}$is the average of the estimated spectra (or periodograms).

This estimator displays some interesting properties. We will show that it coincides with a time domain GMM estimator, with the choice of ϖ corresponding to the choice of a GMM weighting matrix. First we need to define some notation.

Set W be the Fourier matrix. It has typical element $[\mathrm{W}]_{ij} = \frac{1}{\sqrt{2\pi T}}[e^{i\omega_{i-1}j}]$, note that this matrix has the property that $W^{\dagger}W = \mathrm{I}$, where $W^{\dagger}$ is the conjugate transpose of W.

Define W*, $[\mathrm{W}^*]_{mj} = \frac{1}{\sqrt{2\pi T}}[\cos(\omega_m j)]k(j)$, where k(j) is a smoothing window so that k(j)=0 for j>$\mathrm{B}_T, m \in \varpi$. So the number of rows of this matrix is equal to the number of frequencies included in ϖ. Call the number of frequencies included in the band M, so that W* is an M x T matrix. Let,

$$\hat{G} = \begin{bmatrix} \frac{1}{T-1}\sum u_t u_{t-1} \\ \frac{1}{T-2}\sum u_t u_{t-2} \\ \ldots \\ \frac{1}{T-(T-1)}\sum u_t u_{t-(T-1)} \end{bmatrix} \tag{6}$$

be the vector of moment conditions implied by the model. Also define,

$$A = \begin{bmatrix} 1-\frac{1}{T} & .. & -\frac{1}{T} \\ .. & 1-\frac{1}{T} & .. \\ -\frac{1}{T} & .. & 1-\frac{1}{T} \end{bmatrix} \tag{7}$$

$$V = W^{*'}A'AW^*.$$

Now we will show that

$$\hat{G}'V\hat{G} = \sum_{\omega \in \varpi} \left(\hat{f}_{u(\theta)}(\omega) - \bar{f}_{u(\theta)} \right)^2.$$

By definition, $\hat{G}'V\hat{G} = \hat{G}'W^{*'}A'AW^*\hat{G} = (AW^*\hat{G})'(AW^*\hat{G})$. But this is simply,

$$\left[\sum_{\tau} \cos(\omega_1 \tau)k(\tau)G_{\tau} \ldots \sum_{\tau} \cos(\omega_M \tau)e^{i\omega_K \tau}k(\tau)G_{\tau} \right] A'A \begin{bmatrix} \sum_{\tau} \cos(\omega_1 \tau)k(\tau)G_{\tau} \\ \ldots \\ \sum_{\tau} \cos(\omega_M \tau)k(\tau)G_{\tau} \end{bmatrix} \tag{8}$$

$$= \sum_{\omega \in \varpi} \left(\hat{f}_{u(\theta)}(\omega) - \bar{f}_{u(\theta)} \right)^2, \tag{9}$$

since multiplication by the matrix A de-means each column. We could equally well choose the smoothing window in W* to be $k(\tau) \equiv 1$ and thus minimize $\sum_{\omega \in \varpi} \left(I_{u(\theta)}(\omega) - \bar{I}_{u(\theta)} \right)^2$,where $I_{u(\theta)}(\omega)$ is the periodogram of the residual.

This "spectral-GMM" estimator has the special property that the moment conditions being imposed are exactly those which correspond to requiring that $u_t \sim wn$. This estimator allows for whitening of the residuals and, because it is a GSE, it allows for band restricted whitening. This estimator also has a particularly simple form.

2.3 Other Frequency Domain Estimators

In this section, we will now present the GSE cost functions which give rise to the least squares and band restricted least squares estimators, and then (asymptotically) maximum likelihood and band restricted maximum likelihood estimators.

First, we can implement least squares by choosing the trivial cost function, which yields

$$\hat{\theta} = \arg\min_{\theta} \sum_{\omega} I_{u(\theta)}(\omega), \tag{10}$$

where $I_{u(\theta)}(\omega)$ is the periodogram of the residual u_t. To see that this estimator coincides with least squares, write the nonlinear least squares estimator,

$$\hat{\theta} = \arg\min_{\theta} U'U, \tag{11}$$

where U is the Tx1 vector of residuals u_{1t}. Now $U'U = U'W^{\dagger}WU = (WU)^{\dagger}WU$

$$= \sum_{\omega} I_{u(\theta)}(\omega). \tag{12}$$

Band spectral regression is nested within this estimator. Instead of taking the summation over T/2 frequencies, we sum only over the frequencies of interest,

$$\hat{\theta} = \arg\min_{\theta} \sum_{\omega \in \varpi} I_{u(\theta)}(\omega). \tag{13}$$

If, for example, we would like to focus on fluctuations of business cycle frequency we might restrict $\varpi = \left[\frac{\pi}{16}, \frac{\pi}{4} \right]$. Using quarterly data, this band would isolate cycles of length between 2 and 8 years.

Now consider Whittle likelihood estimation,

$$\hat{\theta} = \arg\min_{\theta} \sum_{\omega} \frac{I_{u(\theta)}(\omega)}{\hat{f}_{u(\theta)}(\omega)}$$

where $\hat{f}_{u(\theta)}(\omega)$ is again, the smoothed spectral density of u_t. Under normality and circularity of the residual this estimator can be derived from the familiar time domain likelihood function (see, for example, Harvey (1989)).[4] Diebold, Ohanian, and Berkowitz (1995) advocate band-restricted maximum likelihood, which differs from maximum likelihood by taking the summation over only those frequencies of foremost interest,

$$\hat{\theta} = \arg\min_{\theta} \sum_{\omega \in \varpi} \frac{I_{u(\theta)}(\omega)}{\hat{f}_{u(\theta)}(\omega)}. \tag{14}$$

2.4 Consistency

This section delineates sufficient conditions for the consistency of the GSE under correct specification of the model. Although correct specification is surely an unrealistic assumption, it is of obvious interest to verify that our estimation procedure would be asymptotically valid if we did, in fact, know the true model.

As above, let $\hat{\theta}_T = \arg\min_{\theta} \sum_{\omega \in \varpi} C\left(\hat{f}_{u(\theta)}(\omega)\right)$.

Assumption B1: $\theta \in \Theta$, a compact subset of R^k.

Assumption B2: $g(\cdot, \theta)$ is Borel measurable for each $\theta \in \Theta$ and $g(y_t, \cdot)$ is continuous, uniformly in y_t.

Assumption B3: $\mathrm{E}(g(y_t, \theta)) = 0$ and $g(y_t, \theta)$ has a finite spectral density, L_1-continuous on $\theta \in \Theta$.

Assumption B4: $\mathrm{C}(\cdot)$ is continuous. For both spectral-GMM and the CVM estimator, this assumption is easily verified. In the spectral-GMM case, $\mathrm{C}(\cdot)$ is the quadratic. For the CVM estimator, $\mathrm{C}(\cdot)$ is a compound function, a quadratic operating on a summation.

Assumption B5: There exists a unique θ_0 such that $\mathrm{g}(\mathrm{y}_t, \theta_0) = u(\theta_0) \sim WN$. There must be a unique parameter vector for which the Euler residual is white noise. This is not restrictive in the context of dynamic models. However, some time series models such as ARCH models must be handled with care. Consider for example,

[4] A process is said to be circular if its autocovariance matrix has the form of a circulant. Letting $\gamma(\tau)$ denote the autocovariance at lag τ, a circulant has the property that $\gamma(\tau) = \gamma(T - \tau)$, for $\tau = 1, \ldots T - 1$. Circularity does not hold in infinite moving average processes, in general, but even without circularity Whittle's derivation holds asymptotically.

an AR(1) with ARCH innovations. Even under correct specification, the innovations are not uniquely white noise for any set of ARCH parameters. The parameters may be estimated, however, by noting that the squared innovations have a conditional homoskedastic ARMA representation.

Assumption B6: The loss function , $C(\cdot)$, must have the property that $\arg\min_{\theta} \int_{\varpi} C\left(f_{u(\theta)}(\omega)\right) d\omega = \theta_0$ with $0 < \int_{\varpi} C\left(f_{u(\theta)}(\omega)\right) d\omega$. This says that, given that the model is identified (assumption B4), the loss function must be minimized at the population parameter vector θ_0. This is obviously satisfied for a wide variety of functions, $C(\cdot)$. To make this concrete we will check this assumption for the two Whitening Estimators introduced above.

For the CVM estimator, this condition is trivially fulfilled since

$$\int_{\varpi} C\left(f_{u(\theta)}(\omega)\right) d\omega = \int_{\varpi} \left(\pi F_{u(\theta)}(\omega) - \omega\right)^2 d\omega,$$

which achieves a minimum at $\pi F_{u(\theta)}(\omega) = \omega$, uniquely at θ_0.

Next consider the spectral-GMM. We have $\int_{\varpi} \left(f_{u(\theta)}(\omega) - \bar{f}_{u(\theta)}\right)^2 d\omega$. This is clearly minimized at any θ for which $f_{u(\theta)}(\omega) = k$. Together with assumption B4, θ_0 is the unique value for which this is true.

Theorem 1 *Under the assumptions B1-B6, $\overset{\wedge}{\theta}_T \xrightarrow{p} \theta_0$. Given the regularity conditions discussed above, the generalized spectrum estimator is consistent for the true parameter vector.*

Proof. From assumption B4 and the continuity of C$(\cdot)$, $\int C\left(f_{g(y_t,\theta)}(\omega)\right) d\omega$ is continuous in θ.

Next, note that for fixed θ, we can calculate a consistent estimate of the spectrum, $\hat{f}_{g(y_t,\theta)}(\omega) \xrightarrow{p} f_{g(y_t,\theta)}(\omega)$ uniformly in ω. The $\hat{f}_{g(y_t,\theta)}(\omega_i)$ are asymptotically independent with $\text{cov}\left(\hat{f}_{g(y_t,\theta)}(\omega_i), \hat{f}_{g(y_t,\theta)}(\omega_j\right) = O(T^{-1})$, by, for example, Brillinger (1981). This, in turn, implies that,

$$T^{-1} \sum C\left(\hat{f}_{g(y_t,\theta)}(\omega)\right) \xrightarrow{p} \int C\left(f_{g(y_t,\theta)}(\omega)\right) d\omega.$$

Now, since Θ is compact, we can write

$$\underset{\theta}{sup} \left| T^{-1} \sum C\left(\hat{f}_{g(y_t,\theta)}(\omega)\right) - \int C\left(f_{g(y_t,\theta)}(\omega)\right) d\omega \right| \xrightarrow{p} 0. \qquad (15)$$

Assumption B3 implies that $\hat{f}_{g(y_t,\theta)}(\omega)$ is measurable in y and with continuous C($\cdot$), $C\left(\hat{f}_{g(y_t,\theta)}(\omega)\right)$ is y-measurable. Define N(θ_0) as an open neighborhood around θ_0, and $\overline{N(\theta_0)}$ as its complement. Then $\overline{N(\theta_0)} \cap \Theta$ is compact and $\underset{\theta\in\overline{N(\theta_0)}}{min} \int C\left(f_{g(y_t,\theta)}(\omega)\right) d\omega$ exists. Further define,

$$\eta = \underset{\theta\in\overline{N(\theta_0)}}{min} \int C\left(f_{g(y_t,\theta)}(\omega)\right) d\omega - \int C\left(f_{g(y_t,\theta_0)}(\omega)\right) d\omega.$$

Let A_T be the event $\left|\int C\left(f_{g(y_t,\theta)}(\omega)\right) d\omega - T^{-1}\sum C\left(\hat{f}_{g(y_t,\theta)}(\omega)\right)\right| < \eta/2, \forall\theta$. Then by rearranging terms, A_T implies

$$\int C\left(f_{g(y_t,\theta_T)}(\omega)\right) d\omega < T^{-1}\sum C\left(f_{g(y_t,\theta_T)}(\omega)\right) d\omega - \eta/2. \tag{16}$$

A_T also implies,

$$T^{-1}\sum C\left(\hat{f}_{g(y_t,\theta_0)}(\omega)\right) < \int C\left(f_{g(y_t,\theta_0)}(\omega)\right) d\omega - \eta/2. \tag{17}$$

Now

$$T^{-1}\sum C\left(\hat{f}_{g(y_t,\theta_T)}(\omega)\right) < T^{-1}\sum C\left(\hat{f}_{g(y_t,\theta_0)}(\omega)\right), \tag{18}$$

by definition of θ_T. Combine equation 16 with 18, so that,

$\mathrm{A}_T \Rightarrow$

$$\int C\left(f_{g(y_t,\theta_T)}(\omega)\right) d\omega < T^{-1}\sum C\left(\hat{f}_{g(y_t,\theta_0)}(\omega)\right) - \eta/2. \tag{19}$$

Now add, 17 and 19, which leaves,

$\int C\left(f_{g(y_t,\theta_T)}(\omega)\right) d\omega < \int C\left(f_{g(y_t,\theta_0)}(\omega)\right) d\omega - \eta$. So that when A_T is true, $\theta_T \notin \overline{N(\theta_0)} \cap \Theta$, or $\theta_T \in$N(θ_0).

Now, since, $\underset{T\to\infty}{lim}\ P(\mathrm{A}_T) = 1$,by equation 15, we conclude that $\overset{\wedge}{\theta}_T \overset{p}{\to} \theta_0$.

Q.E.D.

2.5 Asymptotic Normality

Under two additional assumptions, it is possible to show that the Generalized Spectrum estimator is asymptotically Normally distributed. The asymptotic variance will depend not only on the underlying data generating process, but also on the choice of $C(\cdot)$.

Assumption C1: C($\cdot$) is twice continuously differentiable.

Assumption C2: $E\left(C'(\chi_\omega)\chi_\omega\right)^2 < \infty$, χ_ω is an exponentially distributed random variable.

Theorem 2 *Under the additional assumptions C1 and C2,* $\sqrt{T}\left(\hat{\theta}_{GSE}-\theta_0\right) \xrightarrow{d} N(0, H(\theta_0)^{-1}\Sigma H(\theta_0)^{-1})$, *where* $H(\theta_0)$ *is the Hessian of the GSE minimand.*

Proof. Define $Q_T(\theta) = \sum_{\omega} C(\hat{f}(\omega))d\omega$, the quantity being minimized. We have that $\frac{\partial^2}{\partial\theta^2}Q_T(\theta)$ exists and is continuous. It is established in the proof of Theorem 1 that

$$T^{-1}Q_T(\theta) \xrightarrow{P} Q(\theta), \text{ uniformly in } \theta.$$

It is thus straightforward to show that

$$T^{-1}\frac{\partial^2}{\partial\theta^2}Q_T(\theta) \mid_{\theta_T} \xrightarrow{P} \frac{\partial^2}{\partial\theta^2}Q(\theta_0). \tag{20}$$

Next we consider the quantity $\mathrm{T}^{-1/2}\frac{\partial}{\partial\theta}Q_T(\theta) \mid_{\theta_0}$. From the definition of the GSE,

$$\frac{\partial}{\partial\theta}Q_T(\theta) = \sum_{\omega} C'(\hat{f}_\theta(\omega))\frac{\partial\hat{f}_\theta(\omega)}{\partial\theta}. \tag{21}$$

We will evaluate this quantity at the point θ_0, $\sum_{\omega} C'(\hat{f}_{\theta_0}(\omega))\frac{\partial\hat{f}_\theta(\omega)}{\partial\theta}\Big|_{\theta_0}$. The $\hat{f}_{\theta_0}(\omega_i)$, $i =$ 1,.. T/2, form an asymptotically independent triangular array, with cov $\left(\hat{f}_{\theta_0}(\omega_i), \hat{f}_{\theta_0}(\omega_j)\right) = O(T^{-1})$. We may now apply a Central Limit Theorem due to McLeish (1975),

$$T^{-1/2}\sum_{\omega_T} C'(\hat{f}_{\theta_0}(\omega))\frac{\partial\hat{f}_\theta(\omega)}{\partial\theta}\Big|_{\theta_0} \xrightarrow{d} N(0,\Sigma), \tag{22}$$

$\Sigma = \int C'(f_{\theta_0}(\omega))^2 \left(\frac{\partial f_\theta(\omega)}{\partial\theta}\Big|_{\theta_0}\right)^2 d\omega.$

Combining 20 and 22 will be sufficient to prove the theorem by applying the mean value theorem,

$$\frac{\partial Q_T}{\partial\theta}\Big|_{\theta_T} = \frac{\partial Q_T}{\partial\theta}\Big|_{\theta_0} + \frac{\partial^2 Q_T}{\partial\theta^2}\Big|_{\theta^*}(\theta_T-\theta_0), \theta^* \in (\hat{\theta},\theta_0).$$

$$\Longrightarrow \sqrt{T}(\theta_T-\theta_0) = -\left[T^{-1}\frac{\partial^2 Q_T}{\partial\theta^2}\Big|_{\theta^*}\right]^+ T^{-1/2}\frac{\partial Q_T}{\partial\theta}\Big|_{\theta_0}$$

But, plim $\left[T^{-1}\frac{\partial^2 Q_T}{\partial\theta^2}\Big|_{\theta^*}\right]^+ = \frac{\partial^2}{\partial\theta^2}Q(\theta_0)$ by 20. Setting $\mathrm{H}(\theta_0) = \frac{\partial^2}{\partial\theta^2}Q(\theta_0)$, we conclude that

$$\sqrt{T}(\theta_T-\theta_0) \xrightarrow{d} N(0, H(\theta_0)^{-1}\Sigma H(\theta_0)^{-1}).$$

3 Simulation Experiment

In order to illustrate the behavior of GSE, this section reports the results of a simple Monte Carlo experiment. The experiment is designed to mimic some of the characteristics of daily NYSE stock returns data. There is a large literature which studies the effect of high frequency noise on observed stock returns. The measurement error arises primarily from bid-ask spreads and asynchronous trading (e.g., Blume and Stambaugh (1983), Roll (1984)). Consider the following data generating process,

$$x_t = \rho x_{t-1} + \varepsilon_t$$

$$y_t = x_t + u_t.$$

The researcher can only observe y_t, which includes high frequency measurement error u_t. The researcher does not the data generating process for u_t. We are interested in estimating the unknown parameter ρ. We can proceed by using standard ARMA information criteria and estimation tools to fit an ARMA to y_t. Alternatively, we can fit an AR(1) to y_t and omit high frequencies by GSE.

Since we are not interested in the high frequency dynamics in u_t, we will use spectral-GMM with a low frequency band. Intuitively, the spectrum of y_t will be the sum of a low frequency component, x_t and a high frequency component, u_t. By ignoring the high frequencies, we can effectively estimate the spectrum of the low frequency component and thus invert out the persistence parameter.

Following Roll (1984) u_t is modeled as an MA(1) with variance equal to 10% of the variance of y_t. For the purposes of the experiment, ε_t and u_t are taken to be normally distributed. Table 1 reports spectral-GMM estimates for ρ using the low frequency half of available spectra, $\omega \in [0, \frac{\pi}{2}]$. Row 1 contains OLS estimates of ρ, with Monte Carlo standard errors in Row 2. OLS is severely downward biased due to the measurement error. The poor performance of OLS is not surprising, since the model is, in effect, misspecified. Rows 3 and 4 contain estimates which are constructed by fitting an ARMA model to the data using the Hannan and Rissanen (1982) information criterion to select the lag lengths and estimating the parameters via Maximum likelihood. This procedure delivers consistent $\hat{\rho}$, since the order of the ARMA is asymptotically correctly selected. Indeed, the bias of these estimates disappears quickly in the sample sizes studied here. Nevertheless, Spectral-GMM estimates, displayed in Rows 5 and 6, clearly outperform the standard ARMA approach in this experiment. Comparison of Rows 3 and 5 indicates that the GSE has both smaller bias and less Monte Carlo variation, in every sample size.

4 Conclusion

This paper suggests some new techniques for estimating dynamic rational expectations models which are explicity designed to match only subsets of fluctuations in observed data. For example, the methodology allow models to be estimated in the presence of high frequency noise. Generalized Spectral estimators may also be of use in estimating business cycle models or long term growth models because of their inherent focus on subsets of frequencies.

A simple Monte Carlo experiment studies the ability of the GSE to estimate an autoregressive parameter, when the data of interest is observed with high frequency measurement error. It is assumed that the particular form of the measurement error is unkown to the economist. The results of the experiment suggest that the ability of the GSE to "ignore" high frequencies leads to far more precise estimation than standard ARMA techniques in finite samples.

References

[1] Amemiya, T., *Advanced Econometrics*, Cambridge: Harvard University Press, 1985.

[2] Berkowitz, J., "An asymptotically maximum likelihood estimator in the frequency domain," Mimeo, Department of Economics, University of Pennsylvania, 1994.

[3] Blume, M.E. and R.F. Stambaugh, "Biases in Computed Returns: An Application to the Size Effect," *Journal of Financial Economics*, 12(1983), 387-404.

[4] Brillinger, D. R. *Time Series: Data Analysis and Theory* , New York: Holden Day, 1981.

[5] Den Haan, W. J. and A. Marcet, "Accuracy in Simulations," *Review of Economic Studies*, 61(1994), 3-17.

[6] Diebold, F. X., L.E. Ohanian, J. Berkowitz, "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," NBER Technical Working Paper No. 174, 1995.

[7] Durlauf, S. N., "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, 50(1991), 355-376.

[8] Engle, R. F., "Band Spectrum Regression," *International Economic Review*, 15(1974), 1-11.

[9] Engle, R. F., "Exact Maximum Likelihood Methods for Dynamic Regressions and Band Spectrum Regressions," *International Economic Review,* 21(1980), 391-407.

[10] Granger, C.W.J. and P. Newbold, *Forecasting Economic Time Series*, San Diego: Academic Press, 1986.

[11] Hannan, E. J. and J. Rissannen, "Recursive Estimation of Mixed Autoregressive-Moving Average order," *Biometrika,* 69(1982), 81-94.

[12] Hansen, L. P., "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50(1982), 1029-1055.

[13] Roll, Richard "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance,* 39(1984), 1127-1139.

[14] Uhlig, H. and Taylor, J. B., "Solving Non-linear Stochastic Growth Models: A Comparison of Alternative Solution Methods," *Journal of Business and Economic Statistics*, 8(1990), 1-17.

Table 1. Simulation Results

Sample Size	25	50	75	100	150	200	250	1000
OLS	.236	.386	.446	.482	.516	.536	.547	.578
	(.284)	(.212)	(.177)	(.154)	(.128)	(.107)	(.097)	(.047)
ARMA	.333	.664	.770	.815	.844	.861	.870	.894
	(.461)	(.316)	(.216)	(.166)	(.115)	(.085)	(.069)	(.023)
GSE	.787	.838	.856	.856	.875	.882	.886	.896
	(.233)	(.118)	(.084)	(.069)	(.054)	(.043)	(.038)	(.018)

Notes: Alternative estimates of ρ averaged across 2000 Monte Carlo trials. Monte Carlo standard errors are in parentheses. Rows 1-2 are OLS estimates. For Rows 3-4, the order of the moving average is selected by Hannan and Rissanen (1982) information criterion, the parameters are then estimated by MLE. Rows 5-6 are spectral-GMM estimates with $\omega \in [0, \frac{\pi}{2}]$.