Studies in Nonlinear Dynamics & Econometrics

Volume 11, *Issue* 4	2007	*Article* 4

Wavelet Instruments for Efficiency Gains in Generalized Method of Moment Models

Antonis Michis[*] Theofanis Sapatinas[†]

[*]ACNielsen, Antonis.Michis@Nielsen.com
[†]University of Cyprus, t.sapatinas@ucy.ac.cy

Wavelet Instruments for Efficiency Gains in Generalized Method of Moment Models*

Antonis Michis and Theofanis Sapatinas

Abstract

We propose a simple computational method in the context of generalized method of moments for improving the efficiency of regression coefficient estimates. The gains in efficiency arise by incorporating additional moment conditions in the estimation framework based on maximal overlap wavelet packet transforms of the continuous explanatory variables. A major advantage of the proposed method is that it does not require additional exogenous auxiliary information but relies on wavelet packet transforms of the existing continuous explanatory variables. Based on existing theory, we provide theoretical arguments for the proposed methodology, for both linear and nonlinear models, and demonstrate its advantages with both an empirical application concerning two brand demand models and a Monte Carlo simulation study.

*The authors would like to thank Professors Stuart Barber and Guy Nason for some help with the WaveThresh 3.0 software, and an anonymous referee for helpful comments and suggestions.

1 Introduction

In recent years a number of methods have been proposed in the econometrics literature for improving the efficiency of regression coefficient estimates, particularly in the presence of auxiliary information. The theoretical foundations provided by White (1980) and Hansens (1982) have been utilized from a variety of perspectives and several methods have been proposed, primarily in the context of generalized method of moments (GMM), that can lead to efficiency gains, making effective use of the auxiliary information. For instance, Cragg (1983) proposed creating additional auxiliary variables in the form of polynomial functions of the explanatory variables and utilize them as instrumental variables in a two-stage least-squares (2SLS) procedure. When the variance of the model's error term depends on these auxiliary variables, the 2SLS procedure can provide gains in efficiency. More recent examples include information theoretic alternatives to GMM. In this direction, Imbens (1997) and Imbens, Spady & Johnsons (1998) estimate the parameters of interest jointly with the distribution of the data utilizing over-identifying moment conditions. Unlike the empirical distribution case, new and non-equal weights are attached to each observation and the weighted version of the sample moment conditions is set to zero. This approach has also been extended to the case when the sample is not a random draw from the population; see, e.g., Nevo (2002, 2003), Hellerstein & Imbens (1999). In another example, Qian & Schmidt (1999) make use of additional moment conditions that do not depend on the unknown parameters in order to improve efficiency.

In addition to Cragg's polynomial instruments, several other non-linear instrumental variables have been proposed in the literature. Dagenais & Dagenais (1995) suggested a higher moment estimator by utilizing instrumental variables constructed from higher moment functions of the existing exogenous variables. The resulting estimator provides consistent parameter estimates in cases when there is measurement error as well as efficiency gains compared with the ordinary least-squares (OLS) method. A similar approach was also followed by Lewbel (1997) who constructed instrumental variables by utilizing functions of the exogenous variables, with finite third own and cross moments. In addition to dealing with measurement error problems, the instrumental variables were shown to provide gains in efficiency. Donald, Imbens & Newey (2003) suggested the construction of non-linear instrumental variables by employing approximating functions of exogenous variables. Candidate functions include splines, power series, Fourier series and exponentials. Utilizing such instruments, efficient empirical likelihood, GMM and instrumental variable estimators under conditional moment restrictions can be constructed from a sequence of unconditional restrictions. We augment the class of non-linear instrumental variables with further non-linear transformations in the form of wavelets. We achieve this by performing wavelet transformations on the realized values of the exogenous variables which we incorporate as additional instruments in the GMM estimation procedure. By definition, such transformations are deterministic and can hence be considered as part of the information set associated with a data generating process (DGP). As demonstrated by Ramsey & Zhang (1997) and Ramsey & Lampard (1998a, 1998b) wavelet transformations can uncover useful economic information regarding the frequency variation of the economic variables in time. This makes the proposed wavelet instrumental variables particularly attractive for the estimation of relations between the economic time series variables since the additional information generated by the wavelet transformations can lead to important efficiency gains. Our focus is on the maximal overlap discrete wavelet packet transform (MODWPT) which possesses some desirable properties that enable the construction of additional instruments. Wavelet packets form an organized but extremely flexible class of functions of which wavelets are a subset. Wavelets are a modern and powerful mathematical tool and the subject of extensive research currently in statistics, most notably in the fields of non-parametric estimation and time series analysis. Introductions from a statistical perspective can be found in the books by Percival & Walden (2000) and Vidakovic (1999), and in the review articles by Abramovich, Bailey & Sapatinas (2000) and Antoniadis, Bigot, & Sapatinas (2001).

The article is organized as follows. Section 2 provides the necessary wavelet theoretical foundations for the construction of the wavelet instruments. Section 3 provides basic theoretical results for efficiency gains when incorporating additional wavelet instruments into the GMM estimation framework. Results are provided for both linear and non-linear models. Section 4 presents two empirical applications based on the proposed methods that concern one linear and one non-linear brand demand model from a fast moving consumer goods product category of the Canadian retail trade. Section 5 includes a detailed Monte Carlo simulation study in order to investigate the asymptotic gains in efficiency with the use of wavelet instruments. Concluding remarks are made in Section 6.

2 The Construction of Wavelet Instruments

2.1 Wavelets and the Time-Frequency Plane

Representations of time series can be performed either in the time or in the frequency domain. While the first representation is purely concentrated in time, the second is purely concentrated in the frequency domain, summarizing the information available in the time series as a function of frequency. The frequency representation can be obtained by applying the Fourier transform to the observed time series which entails approximating them through a linear combination of sines and cosines (an orthonormal basis). However, Fourier analysis is mostly fruitful when working with stationary time series (see, e.g., Priestley, 1981).

Wavelet analysis offers the opportunity for simultaneous representation of time series in both the time and in the frequency domain. In other words, it represents another, but fixed, tiling of the time-frequency plane. This two-dimensional plane represents time along the horizontal axis and frequency along the vertical axis. Waveforms (segments of time series) can be schematically represented by areas in the time-frequency plane with their width indicating duration and height indicating frequency bandwidth (see, e.g., Figure 4 in Nason & Sapatinas, 2002).

Wavelet analysis utilizes some new families of orthonormal bases termed wavelets. For an appropriate basic function, say ψ, termed the mother wavelet, a set of wavelets, $\psi_{j,k}$, can be obtained by dilating (expansion in the range by a multiplicative factor) and translating (shift in the range by $2^j k$ units) the mother wavelet ψ as follows:

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k), \quad j, k \in \mathbb{Z},$$

where $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is the set of integers. These translations and dilations of the mother wavelet enable the wavelet transform to capture futures in the original times series that are local in both the time and the frequency domain. Through the scaling and translation operations, one is able to analyze the observed time series at different points in time and different frequencies, hence, permitting the analysis of time-varying phenomena, such as non-stationary time series, frequently encountered in many disciplines. (see, e.g., Nason & von Sachs, 1999).

The derived wavelets form an orthonormal basis for $L^2(\mathbb{R})$, the space of functions with finite "energy" on the real line $\mathbb{R} = (-\infty, \infty)$. Having available wavelet bases, one can represent any function $g \in L^2(\mathbb{R})$ as linear combinations of wavelets much like the Fourier approximation, i.e.,

$$g(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} q_{j,k}\psi_{j,k}(t), \tag{1}$$

where

$$q_{j,k} = \int_{\mathbb{R}} g(t)\psi_{j,k}(t);$$

here $q_{j,k}$ forms the series of wavelet coefficients of g.

Formula (1) demonstrates that $g(t)$ can be represented by basis functions, $\psi_{jk}(t)$, at different scales proportional to 2^j for integers j, i.e., a multiresolution analysis. The mother wavelet is usually chosen to be a short-duration oscillation and therefore localized both in time (short-duration) and in frequency (because it oscillates). The derived wavelets are scaled and translated (by $2^j k$) versions of the mother wavelet: the scaling and translation operations permit analysis of time series at different times and frequencies (time-frequency analysis). Wavelets tile the time-frequency plane as follows: as j gets smaller the wavelets become finer and finer scale objects, oscillate more quickly, are packed closer together and the corresponding tiles get taller (they cover a wider frequency range) and thinner (their duration is less). As j gets larger the opposite happens (see, e.g., Figure 4 in Nason & Sapatinas, 2002).

2.2 Wavelet Packet Libraries

Wavelet packets constitute a larger class of orthogonal bases of which wavelets are a subset and the wavelet packets are derived as linear combinations of wavelet functions. Wavelet packet libraries are collections of such bases; any orthonormal basis selected from this library is a wavelet packet basis of $L^2(\mathbb{R})$ (see, e.g., Vidakovic, 1999, Section 5.3).

Unlike the DWT, where the basis elements are associated with scaling and dilation of the mother wavelet (see, e.g., Percival & Walden, 2000, p. 59), the basis elements selected from a wavelet packet library are associated with a specific scale, location and frequency in the range $[0, 1/2]$. In the case of discrete time series, the wavelet packet transform computes all possible filtering combinations efficiently utilizing all basis functions in the library. To illustrate this, suppose we are interested in obtaining the discrete wavelet packet transform (DWPT) of an observed time series of length $T = 2^J$, $J \in \mathbb{Z}_+$, utilizing a wavelet family of even length L. Let $(h_0, h_1, \ldots, h_{L-1})$ and $(g_0, g_1, \ldots, g_{L-1})$ be the discrete high-pass and low-pass filters associated with this wavelet family. In obtaining the vectors of wavelet packet coefficients $w_{j,n}$ at different frequency intervals $\lambda_{j,n} = [n/2^{j+1}, (n+1)/2^{j+1}]$, the following operations are performed recursively to obtain the vector elements:

$$w_{j,n,t} = \sum_{l=0}^{L-1} u_l w_{j-1,[n/2],2t+1-l \bmod T_{j-1}}, \quad t = 0, 1, \ldots, T_j - 1, \tag{2}$$

where

$$u_l = \begin{cases} h_l, & \text{if} \quad n \bmod 4 = 0 \text{ or } 3, \\ g_l, & \text{if} \quad n \bmod 4 = 1 \text{ or } 2, \end{cases}$$

and $[\cdot]$ denotes the 'integer part' operator. In practical applications, the DWPT is obtained through a pyramid algorithm which commences with the input vector $w_{0,0}$ to be the observed data (of dyadic length $T = 2^J$, $J \in \mathbb{Z}_+$) and proceeds recursively. The algorithm involves $O(T \log(T))$ operations. In the first stage, the input vector is high and low pass filtered with downsampling by 2 to obtain the $w_{1,0}$ and $w_{1,1}$ vectors of wavelet packet coefficients respectively, each of length $T/2^j$, where $j = 0, 1 \ldots, J$. In the second stage, the operations in (2) are performed on each one of the vectors $w_{1,0}$ and $w_{1,1}$ to produce the second level of wavelet packet coefficient vectors $w_{2,0}$, $w_{2,1}$, $w_{2,2}$ and $w_{2,3}$. The procedure can then be repeated up to J levels, with $J = \log_2(T)$, where in each case the wavelet packet coefficient vectors cover the frequency interval $[0, 1/2]$ and 2^j vectors are generated at each level.

2.3 Wavelet Instruments

Another popular DWPT is the maximal overlap discrete wavelet packet transform (MODWPT), which generates coefficient vectors of length $T = 2^J$, $J \in \mathbb{Z}_+$, at each level of the transform. The procedure is similar to the DWPT but there is no longer a downsampling of the filtering output involved in the pyramid procedure and a rescaled wavelet filter is used in (2). Specifically, in place of the filters h_l and g_l, we now use the filters $\tilde{h}_l = h_l/\sqrt{2}$ and $\tilde{g}_l = g_l/\sqrt{2}$. The observations in the MODWPT coefficient vectors $\tilde{w}_{j,n}$ are computed again using a pyramid type algorithm based on the equations:

$$\tilde{w}_{j,n,t} = \sum_{l=0}^{L-1} \tilde{u}_l \tilde{w}_{j-1,[n/2],t-2^{j-1}l \bmod T}, \quad t = 0, 1, \ldots, T-1,$$

where

$$\tilde{u}_l = \begin{cases} \tilde{h}_l, & \text{if} \quad n \bmod 4 = 0 \text{ or } 3, \\ \tilde{g}_l, & \text{if} \quad n \bmod 4 = 1 \text{ or } 2. \end{cases}$$

Under this transform, $2T-2$ wavelet packets vectors are generated which have, at every scale, equal length T with the original time series.

In the analysis that follows, we shall be interested in utilizing MODWPT of continuous explanatory time series variables as additional instrumental variables (which will be refer to as *wavelet instruments*) in time series regression econometric models. For this to be feasible, the generated wavelet transforms should be of equal length with the original time series and in addition the time location of the generated wavelet coefficients should have a one to one correspondence with the time points in the original time series. The maximal overlap discrete wavelet packet transform described above does not meet this last requirement but it is possible to circularly shift the elements in each one of the MODWPT vectors generated in order to be aligned with the time points in the original explanatory time series; (see, e.g., Percival & Walden, 2000, p. 234; Nason & Sapatinas, 2002). In the subsequent analysis, it is assumed that all the level J MODWPT coefficient vectors have been circularly shifted to align with the original explanatory time series.

2.4 Time-Scale Decompositions of Economic Time Series

Economic time series enclose information about the activities of multiple agents in the economic environment. These activities arise at several different frequencies in time forming cycles of varying length in the economic data. The different cycles typically enclose different characteristics and considerations by the economic agents which reflect economic reasoning and decision making over different time horizons. A classic distinction of cycles in economics analysis comes in the form of short and long run behavior. In firm theory for example several decisions have a short term orientation such as the introduction of an advertising campaign while other decisions involve long run planning and consideration such as the decision to expand operations to another country. Despite their complex structure and the inclusion of multiple cycle components economic time series are only observed at a single sampling rate which results in an aggregation of the different frequencies in the data. As a consequence important information is lost in the estimation of economic relationships. Ramsey & Lampart (1998a, 1998b) recognized this fact and emphasized the ability of the wavelet analysis to generate time-scale decompositions of economic time series over different frequencies, uncovering variation in the series over cycles of varying length. Using MODWT, the authors conducted two empirical studies by estimating different regression relationships between personal income and consumption in the first study and money and gross domestic product in the second study over different time-scales. Their empirical results confirmed that the relationships between the variables examined did vary significantly across time-scales and also that the phase of the relationship varied with the state of the system. This paper proposes an alternative approach for utilizing the time-scale decompositions of wavelet analysis by constructing wavelet instrumental variables though MODWPT and inserting them in the estimation procedure in order to improve efficiency. The extra information contained in the wavelet instruments regarding the frequency activities included the exogenous variables, provides a more detailed account regarding their variation and behavior in time isolating also features that are specific to each time-scale. When inserted into appropriate estimation procedures we show that this extra information permits a more accurate refinement of the causal effect of the exogenous variables on the dependent variable.

Pricing series of brands and products constitute another example of economic time series enclosing frequency variation over different cycles in time. Leeflang & Witting (1992) categorize pricing variation in short-term, medium-term and long-term. Short-term variation covers cyclical variation of 4 weeks duration or less which is most commonly associated with temporary price reductions (price promotions) and the absence of competitive reactions to observed price changes in the market. Medium-term variation ranges from 4 to 13 weeks and arises mainly by regular price changes and competitive reactions. Cycles exceeding 13 weeks in duration define long term variation which is associated with long-term strategic objectives in the market. In the empirical part of section 4 we decompose the pricing series of two brands and construct wavelet instruments which cover variation over different frequencies. These are then inserted in the estimation of brand demand models and are shown to provide important gains in the efficiency of the coefficient estimates since they permit a more detailed account of the price cycles described above to be included in the estimation process.

The wavelet literature on economics continues to grow. Some recent examples include Gencay, Selcuk & Whitcher (2001a, 2001b, 2001c) who demonstrate the benefits of using wavelet analysis in order to evaluate variations in foreign exchange volatility across different time scales. In another application, Gencay, Selcuk & Whitcher (2003, 2005) found strong evidence that the relation between the return of a portfolio and its beta (systematic risk) in the context of a Capital Asset Pricing Model (CAPM) becomes stronger as the wavelet scale increases. A somewhat different but very interesting class of wavelet applications addresses specification testing issues in econometrics. For instance, Lee & Hong, (2001a, 2001b) provide tests for serial correlation of unknown form and ARCH effects, and Fan & Gencay (2007) for unit roots and cointegration.

3 GMM with Wavelet Instruments

3.1 Linear Models

Consider the following DGP of the classical linear regression form:

$$y = X\beta + \epsilon, \quad \mathbb{E}(\epsilon'\epsilon) = \Delta,$$

where y is a $T \times 1$ vector of time observations on the dependent variable, X is the $T \times K_1$ matrix of observations on K_1 exogenous continuous variables, and ϵ is the $T \times 1$ error term vector with covariance matrix of unknown form Δ. The error term is assumed to satisfy the condition:

$$\mathbb{E}(\epsilon \mid \Omega) = 0, \tag{3}$$

where Ω is the information set. Typically, the information set contains all the potential exogenous variables associated with the specific DGP and in addition all deterministic functions of these variables (see, e.g., Davidson & MacKinnon, 2004, p. 216). Consequently, the MODWPT of the continuous exogenous variables in X belong to the information set since they are, by definition, deterministic functions of these variables.

In the analysis that follows, we shall be interested in constructing a $T \times K$ matrix Z, of instrumental variables that, in addition to the K_1 exogenous explanatory variables, will also contain K_2 additional instruments in the form of the MODWPT coefficients of the continuous exogenous variables in X. Equality (3) implies the $N = K = K_1 + K_2$ moment conditions $Z'(y - X\beta) = 0$ which can be solved for the GMM estimator of β. Specifically, the efficient GMM estimator minimizes the following criterion function:

$$Q(\beta, y) = (y - X\beta)'Z(Z'\Delta Z)^{-1}Z'(y - X\beta). \tag{4}$$

Elaborating Z instead of X in the estimation procedure can be shown to provide gains in efficiency subject to the standard orthogonality and identifiability conditions. For consistent estimation of the parameter vector β, it is required that y, X and ϵ are stationary and ergodic random variables and, in addition to condition (3), the following two identification conditions hold (see, e.g., Wooldridge, 2002, p. 93):

(i) $\text{rank}\,\{\mathbb{E}(Z'Z)\} = K$;

(ii) $\text{rank}\,\{\mathbb{E}(Z'X)\} = K_1$.

As discussed in Section 2, applying a MODWPT on the variables in X generates $2T - 2$ wavelet packet coefficient vectors, each one of length T. This introduces $N = K_1 \times (2T - 2)$ additional wavelet instruments in the estimation procedure, and a selection has to be performed according to some rule in order to reduce the dimensionality of the estimation problem so as to satisfy the order condition

$T > N$. This condition is necessary for the rank conditions (i)-(ii). Furthermore, the generated MOWPT coefficient vectors are not orthogonal between them and an orthogonal transformation must be performed on the matrix of wavelet instruments in order to avoid the problem of multicollinearity. Principal components is one transformation that effectively addresses both issues (see, e.g., Johnson & Wichern, 1998, pp. 458–513).

In the present analysis, we are primarily interested in selecting K_2 principal components that will be inserted as additional instruments in the $T \times K$ matrix of instrumental variables Z. The main consideration is to choose K_2 in such a way that captures as much of the variability inherent in the MODWPT coefficients in the $T \times N$ matrix M, but at the same time retains the number of instruments sufficiently small in order to avoid finite sample problems for the GMM estimator. If a large part of the variability inherent in the matrix M of instrumental variables is explained by the first few principal components, we can represent them by the $T \times K_2$ sub-matrix P_1 and retain the rest in the $T \times (N - K_2)$ sub-matrix P_2, where $P = [P_1, P_2]$ is the matrix of principal components of M. Hence, a simple choice of wavelet instruments to be inserted in the estimation problem (4) is the matrix P_1 which together with matrix X form the matrix of all instrumental variables $Z = [Z_X, Z_{P_1}] = [X, P_1]$.

The efficient GMM estimator has the well known form:

$$\hat{\beta}_l = (X'Z\Sigma^{-1}Z'X)^{-1}X'Z\Sigma^{-1}Z'y,$$

where $T^{-1}Z'X \xrightarrow{p} D$ and $T^{-1}Z'\Delta Z \xrightarrow{p} \Sigma$ is the asymptotic covariance matrix of the sample moments. (In what follows, "$\xrightarrow{p}$" denotes convergence in probability.) The asymptotic covariance matrix of the GMM estimator is:

$$\mathbb{V}\text{ar}(\hat{\beta}_l) = (D'\Sigma^{-1}D)^{-1}$$

and can be consistently estimated from:

$$\widehat{\mathbb{V}\text{ar}}(\hat{\beta}_l) = T(X'Z\hat{\Sigma}^{-1}Z'X)^{-1},$$

where $\hat{\Sigma}$ is a heteroskedasticity and autocorrelation consistent estimator of Σ. Such estimators have been proposed by, e.g., Hansen (1982), Newey & West (1987). To illustrate the possible efficiency gains generated by incorporating the additional wavelet instruments in the GMM estimation procedure, partition the matrices D and Σ as follows, where D_{XP_1} is the sub-matrix of elements associated with the wavelet instruments in P_1:

$$D = \begin{bmatrix} D_{XX} \\ D_{XP_1} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XP_1} \\ \Sigma_{P_1X} & \Sigma_{P_1P_1} \end{bmatrix}.$$

Now, consider the estimator:

$$\tilde{\beta}_l = (X'Z_X\Sigma_{XX}^{-1}Z_X'X)^{-1}X'Z_X\Sigma_{XX}^{-1}Z_X'y,$$

where $\tilde{\beta}_l$ does not contain the additional matrix of wavelet instruments P_1. It can be then shown that:

$$\mathbb{V}\text{ar}(\tilde{\beta}_l) \geq \mathbb{V}\text{ar}(\hat{\beta}_l)$$

if and only if

$$D_{XP_1} \geq \Sigma_{P_1X}\Sigma_{XX}^{-1}D_{XX}. \quad (5)$$

For a detailed proof of the relevant theorem concerning efficiency gains arising from utilizing additional instruments see, e.g., Peracchi (2001, Theorem 11.6). When incorporating wavelet instruments as demonstrated above, test statistics and confidence intervals can be constructed in the usual way.

3.2 Non-Linear Models

Let Θ be a compact subset of $\mathbb{R}^{K_1}$ and define the moment functions $\Psi : H \times \Theta \mapsto \mathbb{R}^K$, $K \geq K_1$, based on a DGP ϕ involving a nonlinear regression function. Let $H = [y, Z]$ be a multivariate stationary and ergodic random variable which includes the endogenous variable y and the exogenous variables in Z as defined in Section 3.1. The moment function Ψ is Borel measurable in Z and twice continuously differentiable with respect to $\beta \in \Theta$, the K_1-dimensional parameter vector of interest. In addition, assume:

(i) $\mathbb{E}_\phi[|\Psi(y, Z; \beta)|] < \infty$;

(ii) $T^{-1}\sum_{t=1}^{T}\Psi(y_t, Z_t; \beta) \xrightarrow{p} \mathbb{E}_\phi[\Psi(y, Z; \beta)]$;

(iii) $\mathbb{E}_\phi[\Psi(y, Z; \beta^*)] = 0$ for a unique element $\beta^* \in \Theta$;

(iv) the covariance $\mathbb{E}_\phi[\Psi(y, Z; \beta^*)\Psi(y, Z; \beta^*)'] = \Sigma_0$ and the Jacobian $\mathbb{E}_\phi[\nabla_\beta\Psi(y, Z; \beta)] = \underline{\Psi}$ matrices are of full rank.

In the over-identified case, i.e., when $K \geq K_1$ (where $K = K_1 + K_2$ and K_2 is the number of wavelet instruments as in Section 3.1) the GMM estimator minimizes the following sample analog of the population moment conditions in order to be as close to zero as possible:

$$\hat{\beta} = \arg\min_{\beta\in\Theta}\left[\sum_{i=1}^{T}\Psi(y_i, Z_i; \beta)\right]' \Xi \left[\sum_{t=1}^{T}\Psi(y_i, Z_i; \beta)\right],$$

where Ξ is a $K \times K$ finite, symmetric, positive definite matrix. The asymptotically best GMM estimator corresponds to choosing Σ_0^{-1} for Ξ in which case the asymptotic covariance matrix of the GMM estimator becomes $\mathbb{V}\text{ar}(\hat{\beta}) = T^{-1}(\underline{\Psi}'\Sigma_0\underline{\Psi})^{-1}$. Chamberlain (1987) showed that this estimator is asymptotically efficient in the general class of semi-parametric estimators.

To illustrate the possible efficiency gains generated by incorporating the additional wavelet instruments in the GMM estimation procedure when a nonlinear regression function is involved, we adopt a similar reasoning as for the linear case in deriving relation (5). Let the moment conditions be written in a generalized residual function form:

$$\mathbb{E}_\phi[\Psi(y, Z; \beta^*)] = \mathbb{E}_\phi[Z^* \psi(y, Z; \beta^*)] = 0,$$

where $Z^* = Z\Pi$ and $\Pi = (Z'\Delta_0 Z)^{-1} Z'\underline{\psi}$ is the full rank $K \times K_1$ projection matrix of the optimal linear combinations of the $\overline{K}$ instruments, $\underline{\psi}$ is a $T \times K_1$ Jacobian matrix with elements $(\partial\psi/\partial\beta_i)$ $(i = 1, 2, \ldots, K)$, and $\mathbb{E}_\phi[\psi(y, Z; \beta)\psi(y, Z; \beta)'] = \Delta_0$. The covariance matrix of the asymptotically best GMM estimator then takes the form:

$$\mathbb{V}\mathrm{ar}(\hat{\beta}) = \left(D_0' \Sigma_0^{-1} D_0\right)^{-1},$$

where $T^{-1} Z'\Delta_0 Z \xrightarrow{p} \Sigma_0$ and $T^{-1} Z'\underline{\psi} \xrightarrow{p} D_0$. By replacing Σ_0 with a heteroskedasticity and autocorrelation consistent estimator $\hat{\Sigma}_0$, the covariance matrix of the GMM estimator can be estimated by:

$$\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}) = T\left(\hat{\underline{\psi}}' Z \hat{\Sigma}_0^{-1} Z' \hat{\underline{\psi}}\right)^{-1},$$

where $\hat{\underline{\psi}} \equiv \underline{\psi}(\hat{\beta})$ (see, e.g., Davidson & MacKinnon, 2004, p. 377).

Implicit in the above GMM estimation framework are the order condition $T > K$ and the analog of the rank conditions (i)-(ii) in Section 3.1. Specifically condition (i) remains essentially the same while in place of condition (ii) we now require that $\mathrm{rank}\{\mathbb{E}(Z'\underline{\psi})\} = K_1$ (see, e.g., Wooldridge, 2002, p. 426). Consequently, the discussion of Section 3.1 for effective dimension reduction in the construction of the matrix of wavelet instruments equally applies here.

The following proposition, whose proof is given in the Appendix, is a direct extension to the non-linear case of the arguments providing relation (5).

Proposition 3.1 *Assume the standard GMM conditions (i) – (iv) to hold. Let* $Z = [Z_X, Z_{P_1}] = [X, P_1]$ *be the matrix of instruments and partition the matrices* D_0 *and* Σ_0 *as follows:*

$$D_0 = \begin{bmatrix} D_{X\underline{\psi}} \\ D_{P_1\underline{\psi}} \end{bmatrix} \quad \text{and} \quad \Sigma_0 = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XP_1} \\ \Sigma_{P_1X} & \Sigma_{P_1P_1} \end{bmatrix}.$$

Let also

$$\hat{\beta}_{nl} = \arg\min_{\beta\in\Theta} \left[\sum_{i=1}^{T} Z_i'\psi(y_i, Z_i; \beta)\right]' \Xi \left[\sum_{i=1}^{T} Z_i'\psi(y_i, Z_i; \beta)\right],$$

and

$$\tilde{\beta}_{nl} = \arg\min_{\beta\in\Theta} \left[\sum_{i=1}^{T} X_i'\psi(y_i, X_i; \beta)\right]' \Xi \left[\sum_{i=1}^{T} X_i'\psi(y_i, X_i; \beta)\right],$$

where $T^{-1} Z'\Delta_0 Z \xrightarrow{p} \Sigma_0$ *and* $T^{-1} Z'\underline{\psi} \xrightarrow{p} D_0$*. Then,*

$$\mathbb{V}\mathrm{ar}(\tilde{\beta}_{nl}) \geq \mathbb{V}\mathrm{ar}(\hat{\beta}_{nl})$$

if and only if

$$D_{P_1\underline{\psi}} \geq \Sigma_{P_1X} \Sigma_{XX}^{-1} D_{X\underline{\psi}}.$$

4 An Empirical Application

In this section, we empirically examine the efficiency gains arising by incorporating additional wavelet instruments in the estimation procedure in the context of two brand demand models. The data were supplied by ACNielsen-Canada and concern weekly scanning observations on two leading brands from a well defined fast moving consumer goods category that will be referred to as 'brand 1' and 'brand 2'. Measurements concern a Canadian province as covered by the ACNielsen retail measurement services, hence the sales figures for each brand refer to its total sales volume at the province level. The sales volume is measured in the category's unit of measurement as adopted by ACNielsen, i.e., kgs. There are 117 weekly observations associated with each brand ending in September 2003. Since the data refer to a real life empirical project the names of the brands and the exact dates of the observations (i.e., week of a specific promotional activity) cannot be revealed due to strict confidentiality issues.

For 'brand 1' we examine a linear demand model specification with the variables included in logarithm form,

$$\mathrm{LogSALES}_t = \beta_0+\beta_1\,\mathrm{LogPRICE}_t+\beta_2\,\mathrm{LogCOMPR}_t+\beta_3\,\mathrm{LogPROMO}_t+\beta_4\,\mathrm{LogDISTR}_t.$$

Price refers to the relative average price over the range of stock keeping units (SKU) under the specific brand. COMPR represents similarly defined relative average prices of the main competitive brands in the market. Two such prices were found to have a statistically significant effect on the sales volume of 'brand 1' and were included in the model specification as 'compr 1' and 'compr 2' (this refers to the relative average price of 'brand 2' below). PROMO refers to promotional activities relevant to 'brand 1'. Two kinds of such activities appeared for 'brand 1' in the market: (a) 'features', which is an advertisement in a flyer that comes in the local newspaper where the brand's SKU are shown and (b) 'display', which is any secondary location (over and above their regular store location) of the brand's SKU within a given store. Both of them have been included as binary variables. DISTR refers to the brand's numeric handling distribution. This is the percentage of stores in the province to which any of the brand's SKU are distributed.

For 'brand 2' we examine a non-linear demand model specification with market share as the dependent variable,

$$\mathrm{SHARE}_t = \beta_0 + \exp\left\{\beta_1\,\mathrm{PRICE}_t + \beta_2\,\mathrm{COMPR}_t + \beta_3\,\mathrm{PROMO}_t + \beta_4\,\mathrm{DISTR}_t\right\}.$$

Variable interpretations are similar to the case of 'brand 1', with the exception that one competitive brand price and one promotion type ('features') were incorporated in the estimation of model.

4.1 Wavelet Packet Analysis

The MODWPT was applied on all the continuous variables of each model, followed by principal components analysis on all the derived wavelet packets coefficient vectors. In this way, the desired dimension reduction was achieved and the first 10, 15, 20, 25 and 30 principal components were successively used as additional instruments in the GMM estimation of the linear and non-linear models discussed respectively in Sections 3.1 and 3.2.

Through the wavelet transform the different cycles of periodicity inherent in the pricing and distribution time series are now uncovered and inserted in the estimation of the demand models. As emphasized in section 2.5 this information was not previously observed due to the frequency aggregation in the observed sampling rate of the data. This is particularly true for the pricing series which most commonly enclose short term, medium term and long term cyclical variation as per the analysis of Leeflang & Witting (1992). Using spectral analysis Bronnenberg, Mela & Boulding (2006) examined empirically the periodicity of pricing and verified the existence of several different cycles with distinct characteristics. Among other empirical generalizations the authors concluded that the cross-brand correlation in prices occurs at multiple planning horizons, and the planning horizon of the predominant interaction does not typically coincide with the sampling rate of the data. Furthermore, aggregating pricing interactions across frequencies obscures distinct and different interactions. Such cases are particularly fruitful for wavelet transformations which effectively decompose the data into the underlying frequencies while at the sale time preserving their time location characteristics.

The wavelet packet analysis was performed using the `WaveThresh 3.0` software that is freely available from `http://www.stats.bris.ac.uk/ wavethresh/`. `WaveThresh 3.0` computes the wavelet packet coefficient vectors for time series of length that is a power of two. In order to overcome this restriction, we pad the time series with zeros up to length $T = 128 = 2^7$. As Nason & Sapatinas (2000) point out, due to the time localization of the wavelet transform, the extra zeroes do not affect the majority of the wavelet coefficients except at very coarse scales where they get included in their calculation. This generates 254 wavelet packet coefficient vectors for each variable with 128 coefficients (observations) in each vector. After the appropriate time alignment of the coefficients we remove from each vector the coefficients associated with the padded zeros in order to obtain wavelet packet coefficient vectors of length 117.

The choice of a wavelet basis and its associated filters is another point of consideration when applying the MODWPT. Ideally, the choice of a wavelet basis should be such that it is able to capture as much of the information laying in the analyzed time series as possible. In this respect, it is advisable to experiment with several bases in the context of a specific application in order to arrive at the best possible representations. In deriving the MODWPT of the continuous variables in our empirical application, the Daubechies' least asymmetric wavelet filter of length 10 is employed (see, e.g., Daubechies, 1992, Table 6.3), which was found to provide particular good prediction performance in the context of our empirical model.

4.2 Results

GMM estimation was performed first without the addition of any wavelet instruments (GMM) and then with successive additions of 10 (GMM 10 PC), 15 (GMM 15 PC), 20 (GMM 20 PC), 25 (GMM 25 PC) and 30 (GMM 30 PC) wavelet instruments. In each case, the wavelet instruments referred to the top ranked principal components (PC) from a principal component analysis on the matrix of all derived MODWPT coefficient vectors as explained in Section 3. Table 1 summarizes the differences in the coefficient estimates between the different GMM procedures. In all cases, the coefficient estimates have similar values and meaningful economic interpretation. An exception is the drop in the price coefficient estimate of the non-linear model observed between GMM and the rest 5 GMM estimates utilizing the wavelet instruments.

Table 1 also contains the estimates of the standard errors associated with each estimation procedure and the achieved reductions in the standard errors by the successive use of the additional wavelet instruments. Heteroskedasticity and autocorrelation robust standard errors were obtained through the use of Bartlett weights as suggested by Newey & West (1987) with the lag truncation parameter being 4. As it is evident in the results, there are successive reductions on the standard error estimates of all the coefficients with the use of increasing number of wavelet instruments. For the GMM 30 PC case these reductions range from 34% to 58% for the linear model coefficients and from 58% to 75% for the non-linear model coefficients. We did not proceed further with the addition of more wavelet instruments because after the level of 30 PC evidence of deterioration to the standard errors began to appear something which is in line with existing theoretical knowledge stressing negative effects to the finite sample properties of the GMM estimator when a high number of instruments is incorporated in the estimation procedure. Finally, Table 1 provides values of the chi-square test for over-identifying restrictions which confirm the validity of the additional wavelet moment conditions.

The above mentioned GMM results were then compared with the estimator proposed in Cragg (1983) where the additional instruments introduced in the GMM framework were in the form of polynomial (PL) functions of the continuous explanatory variables, $z_{jt} = x_t^{j+1}$. Successively, one polynomial form from each next continuous explanatory variable was used until achieving the required number of instruments for each estimation round. Table 2 provides the corresponding results for the differences in the coefficient estimates, the standard error reductions and the chi-square tests when using polynomial instruments. As it is evident in the results, the use of polynomial instruments provides significant gains in efficiency which for some coefficients (for example 'distribution' and 'features' in the non-linear model) are higher from those realized when utilizing wavelet instruments. However, this is only true when the overall number of instruments is small. At the level of 30 instruments, wavelets outperform polynomial functions in all the coefficient standard error estimates providing higher reductions up to the level of 20% in the linear case and 30% in the non-linear case.

5 A Monte Carlo Simulation Study

In order to investigate the asymptotic gains in efficiency associated with the use of wavelet instruments, we contacted a detailed Monte Carlo simulation study similar in design to the one proposed by Cragg (1983). The results of the study are summarized in Table 3. This involved an equation with one constant and one explanatory variable of the following form,

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t,$$

where the values of the scalars β_0 and β_1 were obtained using independent, log-normally distributed pseudo-random variables. These were held fixed in the subsequent replications of the experiment. The errors ϵ_t were independent, normally distributed pseudo-random variables with mean zero and variances,

$$\sigma_t^2 = \gamma_0 + \gamma_1 x_t + \gamma_2 x_t^2.$$

In the initial experiment, a sample size of $T = 32$ was generated and the parameter values were set as $\beta = (1.0, 1.0)$ and $\gamma = (0.1, 0.2, 0.3)$ which ensured substantial heteroskedasticity. Each conducted experiment included 1000 replications using independent samples from the specified model. The distributional and functional form specifications of the model are such that all that matters are the relative values of the parameters in γ, which permits one to present all results in relative terms. As noted by Cragg (1983), the estimated error variance $\hat{\Sigma}$ and the bias $(\hat{\beta}^A - \beta)$ of the feasible IV estimator, $\hat{\beta}^A = (X'Z(Z'\hat{\Sigma}Z)^{-1}Z'X)^{-1}X'Z(Z'\hat{\Sigma}Z)^{-1}Z'y$, with variance $\widehat{\mathbb{V}\text{ar}}(\hat{\beta}_A) = (X'Z(Z'\hat{\Sigma}Z)^{-1}Z'X)^{-1}$, do not depend on β and can be selected arbitrarily. The same however is not true for the variables x_t and the specified form of heteroskedasticity which exert influence on the results. The model specification ensured substantial efficiency difference between the least-squares estimator, $\hat{\beta}_L = (X'X)^{-1}X'y$, with variance $\mathbb{V}\text{ar}(\hat{\beta}_L) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$, and the true Aitken estimator, $\tilde{\beta}^E = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$, with variance $\mathbb{V}\text{ar}(\tilde{\beta}^E) = (X'\Sigma^{-1}X)^{-1}$.

For each replication of the experiment, 3 estimated variance values are reported that make use of wavelet instruments, in relative terms to their least-squares population variances. First, the relative asymptotic variances are provided by the diagonal elements of $(X'Z(Z'\Sigma Z)^{-1}Z'X)^{-1}$ relative to those of $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$. Second, the variances of the actual feasible IV estimators relative to least-squares estimator are provided by the ratio $(\hat{\beta}_i^A - \beta_i)^2/\mathbb{V}\text{ar}(\hat{\beta}_L)_{ii}$. Third, the relative estimated variances are provided in which case the numerator is $\widehat{\mathbb{V}\text{ar}}(\hat{\beta}_A)$.

In the first phase of the experiment, estimations were performed with the use of 4 wavelet instruments obtained from each of the 4 levels of coefficients arising from the MODWPT (level 0 contained the higher number of coefficient vectors, 32). In addition, results for 3 different wavelet families are reported in order to observe the impact of the choice of wavelet filter on the results. These were the Daubechies Least Asymmetric (with a filter of length 10), the Daubechies Extremal Phase (with a filter of length 10) and the Haar wavelet family. The findings which are summarized in Table 3 bare similarities with Cragg's polynomial instruments.

There are significant gains in efficiency in the case of the asymptotic and the estimated variances at all levels and with all the wavelet families. The highest gains arise by the use of the Haar family and the level 0 instruments. This could be due to the finer details provided by the wavelets coefficients at this level which provide superior variability and information in the estimation framework. The same was not true however for the actual variances of the slope coefficient, whose variance values were not significantly different from those of least-squares. This could be an indication for the need of a higher number of wavelet instruments until important efficiency gains are realized. In order to investigate this, estimations were performed using 10 and 15 wavelet instruments from the Daubechies least asymmetric MODWPT at levels 0 and 1, where 32 and 16 wavelet coefficient vectors were generated respectively. The results indicate a clear improvement in the efficiency gains arising from the variance of the actual estimator and even higher gains for the asymptotic and estimated variances. The robustness of the efficiency gains arising from the use wavelet instruments were also investigated in the presence of more exogenous variables. The model specification in this case was,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t,$$

where the values of the scalars x_{1t} and x_{2t} were obtained again using independent, log-normally distributed pseudo-random variables which were held fixed in the subsequent replications of the experiment. The errors were independent, normally distributed pseudo-random variables with mean zero and variances,

$$\sigma_t^2 = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{1t}^2 + \gamma_3 x_{2t} + \gamma_4 x_{2t}^2.$$

Similarly with the initial experiment, a sample size of $T = 32$ was generated and the parameter values were set as $\beta = (1.0, 1.0, 2.0)$ and $\gamma = (0.1, 0.2, 0.3, 0.2, 0.5)$ which again ensured substantial heteroskedasticity. Each conducted experimented included 1000 replications using independent samples from the specified model. With the use of 10 wavelet instruments from levels 0 and 1 of the Daubechies Least Asymmetric MODWPT, significant efficiency gains were realized with all the estimators and for all the model coefficients. In a second phase, the effect of the sample size on the realized efficiency gains was examined using the design proposed by Cragg (1983). The experiment in this case included generation of sample sizes of 32, 128 and 512 observations with 1000 replications involved in each case similarly to phase one. The actual and estimated variances relative to their asymptotic values were calculated, arising from the use of 4 wavelets instruments from each level of the Daubechies Least Asymmetric MODWPT in the estimators. The results clearly suggest that the actual and estimated variances do get closer to their asymptotic values as the sample size is increasing.

In summarizing the results of the Monte-Carlo simulation study, we conclude that the MODWPT can provide useful wavelet instruments that can lead to significant gains in efficiency. In our study, we have found evidence that higher gains in efficiency arise by the use of instruments from the finer resolution levels of the

MODWPT and with the use of the Haar filter. Of even more importance is to choose a sufficiently high number of wavelet instruments in order to effectively capture all the ranges of frequency variation hidden in the original data. In addition, we have found evidence that the efficiency gains are robust to the use of additional exogenous variables and the actual and estimated variances do converge to their asymptotic values as the sample size is increasing.

6 Conclusions

We have incorporated wavelets, a modern and powerful mathematical tool, in the context of generalized method of moments estimation framework. As deterministic functions of the realized values of economic variables, maximal overlap wavelet packet transforms provide valid instruments that can be effectively incorporated in the generalized method of moments estimation framework for improving the efficiency of regression coefficient estimates. Based on existing theory, we have provided theoretical arguments for the proposed methodology, for both linear and non-linear models, and have demonstrated its advantages with both an empirical application concerning two brand demand models and a Monte Carlo simulation study.

Appendix: Proof of Proposition 3.1

Following Peracchi (2001, pp. 383–384),

$$\mathbb{V}\mathrm{ar}(\tilde{\beta}_{nl}) - \mathbb{V}\mathrm{ar}(\hat{\beta}_{nl}) = \left(D'_{X\underline{\psi}}\Sigma_{XX}^{-1}D_{X\underline{\psi}}\right)^{-1} - \left(D'_0\Sigma_0^{-1}D_0\right)^{-1}.$$

Consequently,

$$\mathbb{V}\mathrm{ar}(\tilde{\beta}_{nl}) \geq \mathbb{V}\mathrm{ar}(\hat{\beta}_{nl})$$

if and only if

$$D'_0\Sigma_0^{-1}D_0 \geq D'_{X\underline{\psi}}\Sigma_{XX}^{-1}D_{X\underline{\psi}}.$$

Obviously,

$$\begin{aligned} D'_0\Sigma_0^{-1}D_0 &= \begin{bmatrix} D'_{X\underline{\psi}} & D'_{P_1\underline{\psi}} \end{bmatrix} \Sigma_0^{-1} \begin{bmatrix} D'_{X\underline{\psi}} \\ D'_{P_1\underline{\psi}} \end{bmatrix} \\ &= D'_{X\underline{\psi}}\left(\Sigma_{XX}^{-1} + \Sigma_{XX}^{-1}\Sigma_{XP_1}H^{-1}\Sigma_{P_1X}\Sigma_{XX}^{-1}\right)D_{X\underline{\psi}} \\ &\quad -D'_{P_1\underline{\psi}}H^{-1}\Sigma_{P_1X}\Sigma_{XX}^{-1}D_{X\underline{\psi}} \\ &\quad -D'_{XX}\Sigma_{XX}^{-1}\Sigma_{XP_1}H^{-1}D_{P_1\underline{\psi}} + D'_{P_1\underline{\psi}}H^{-1}D_{P_1\underline{\psi}}. \end{aligned}$$

Since Σ_0 is symmetric and positive definite, $H = \Sigma_{P_1P_1} - \Sigma_{P_1X}\Sigma_{XX}^{-1}\Sigma_{XP_1}$ is symmetric and positive definite, implying that

$$\begin{aligned} D_0'\Sigma_0^{-1}D_0 - D_{X\underline{\psi}}'\Sigma_{XX}^{-1}D_{X\underline{\psi}} &= D_{X\underline{\psi}}'\Sigma_{XX}^{-1}\Sigma_{XP_1}H^{-1}\Sigma_{P_1X}\Sigma_{XX}^{-1}D_{X\underline{\psi}} - D_{P_1\underline{\psi}}'H^{-1}\Sigma_{P_1X}\Sigma_{XX}^{-1}D_{X\underline{\psi}} \\ &\quad - D_{XX}'\Sigma_{XX}^{-1}\Sigma_{XP_1}H^{-1}D_{P_1\underline{\psi}} + D_{P_1\underline{\psi}}'H^{-1}D_{P_1\underline{\psi}} \\ &= \left(D_{P_1\underline{\psi}}' - D_{X\underline{\psi}}'\Sigma_{XX}^{-1}\Sigma_{XP_1}\right)H^{-1}\left(D_{P_1\underline{\psi}}' - \Sigma_{P_1X}\Sigma_{XX}^{-1}D_{X\underline{\psi}}\right). \end{aligned}$$

Therefore,

$$\mathbb{V}\mathrm{ar}(\tilde{\beta}_{nl}) \geq \mathbb{V}\mathrm{ar}(\hat{\beta}_{nl})$$

if and only if

$$D_{P_1\underline{\psi}} - \Sigma_{P_1X}\Sigma_{XX}^{-1}D_{X\underline{\psi}} \geq 0.$$

This completes the proof of Proposition 3.1. □

Table 1 GMM Estimation with Wavelet Instruments

Coefficient	*GMM*	*GMM (10 PC)*	*GMM (15 PC)*	*GMM (20 PC)*	*GMM (25 PC)*	*GMM (30 PC)*
Difference in Coefficient Estimates between GMM and GMM with Wavelet instruments						
Linear Model						
Intercept	-	-0.330	-0.242	-0.241	-0.378	-0.268
Log Price	-	-0.013	-0.075	-0.095	-0.052	-0.056
Log CPrice1	-	-0.199	-0.123	-0.153	-0.140	-0.060
Log CPrice2	-	0.163	-0.003	0.053	-0.069	-0.043
Log Dist	-	0.192	0.156	0.152	0.244	0.169
Feature	-	0.008	0.006	0.008	0.005	-0.002
Display	-	-0.003	-0.001	-0.002	-0.004	-0.003
Non-Linear Model						
Intercept	-	1.298	1.324	2.768	2.613	2.407
Price	-	-0.250	-0.305	-0.369	-0.357	-0.298
CPrice1	-	0.138	0.192	0.143	0.140	0.096
Dist	-	0.000	-0.001	0.000	0.000	0.000
Feature	-	-0.015	-0.015	0.006	0.010	0.012
Coefficient Standard Errors with Additional Wavelet Instruments						
Linear Model						
Intercept	1.217	0.948	0.937	0.874	0.758	0.662
Log Price	0.352	0.269	0.248	0.188	0.187	0.149
Log CPrice1	0.554	0.424	0.381	0.370	0.331	0.317
Log CPrice2	0.546	0.457	0.435	0.401	0.351	0.285
Log Dist	0.695	0.548	0.540	0.504	0.429	0.376
Feature	0.025	0.018	0.017	0.015	0.012	0.010
Display	0.010	0.009	0.008	0.007	0.007	0.007
Non-Linear Model						
Intercept	3.488	2.450	1.939	1.477	1.332	1.222
Price	0.557	0.264	0.193	0.158	0.159	0.135
CPrice1	0.304	0.171	0.149	0.133	0.129	0.101
Dist	0.002	0.001	0.001	0.001	0.001	0.001
Feature	0.035	0.025	0.023	0.015	0.013	0.012
Chi-Square Test for Wavelet Overidentifying Restrictions						
Linear model	-	8.6643	11.304	12.2736	14.3657	15.625
Non-Linear Model	-	12.6914	13.5804	16.4283	18.0188	18.3478
Standard Error Reductions with Additional Wavelet Instruments						
Linear Model						
Intercept	-	22.05%	22.94%	28.14%	37.70%	45.56%
Log Price	-	23.63%	29.65%	46.55%	46.83%	57.65%
Log CPrice1	-	23.52%	31.18%	33.18%	40.31%	42.86%
Log CPrice2	-	16.15%	20.20%	26.53%	35.66%	47.74%
Log Dist	-	21.11%	22.20%	27.44%	38.18%	45.85%
Feature	-	25.61%	30.08%	40.65%	50.41%	58.54%
Display	-	12.40%	23.20%	26.60%	28.70%	34.80%
Non-Linear Model						
Intercept	-	29.76%	44.42%	57.66%	61.82%	64.97%
Price	-	52.62%	65.40%	71.61%	71.42%	75.71%
CPrice1	-	43.92%	51.08%	56.25%	57.66%	66.80%
Dist	-	21.99%	27.75%	45.03%	49.06%	58.74%
Feature	-	28.41%	33.81%	57.39%	63.35%	66.19%

Table 2 GMM Estimation with Polynomial Instruments

Coefficient	*GMM*	*GMM (10 PC)*	*GMM (15 PC)*	*GMM (20 PC)*	*GMM (25 PC)*	*GMM (30 PC)*
Difference in Coefficient Estimates between GMM and GMM with Polynomial instruments						
Linear Model						
Intercept	-	-0.242	-0.073	-0.092	-0.108	-0.058
Log Price	-	-0.239	-0.210	-0.228	-0.116	-0.085
Log CPrice1	-	0.325	0.363	0.393	0.063	0.156
Log CPrice2	-	-0.052	-0.090	-0.058	0.034	-0.199
Log Dist	-	0.120	0.023	0.026	0.057	0.048
Feature	-	-0.021	-0.020	-0.021	-0.006	-0.005
Display	-	-0.001	-0.001	-0.001	-0.005	-0.005
Non-Linear Model						
Intercept	-	2.757	2.586	1.999	1.990	1.684
Price	-	-0.407	-0.348	-0.316	-0.308	-0.251
CPrice1	-	0.176	0.131	0.140	0.135	0.103
Dist	-	0.000	0.000	0.000	0.000	0.000
Feature	-	-0.008	-0.009	-0.008	-0.008	-0.008
Coefficient Standard Errors with Additional Polynomial Instruments						
Linear Model						
Intercept	1.217	1.001	0.884	0.871	0.802	0.829
Log Price	0.352	0.284	0.258	0.251	0.227	0.225
Log CPrice1	0.554	0.451	0.432	0.421	0.358	0.345
Log CPrice2	0.546	0.469	0.466	0.460	0.456	0.405
Log Dist	0.695	0.560	0.493	0.487	0.448	0.466
Feature	0.025	0.016	0.015	0.015	0.011	0.011
Display	0.010	0.009	0.009	0.009	0.009	0.009
Non-Linear Model						
Intercept	3.488	2.812	2.298	2.108	1.919	1.815
Price	0.557	0.289	0.227	0.231	0.203	0.190
CPrice1	0.304	0.178	0.130	0.131	0.124	0.118
Dist	0.002	0.001	0.001	0.001	0.001	0.001
Feature	0.035	0.023	0.021	0.022	0.022	0.023
Chi-Square Test for Polynomial Overidentifying Restrictions						
Linear model	-	4.4097	4.6763	4.8536	7.4199	8.1794
Non-Linear Model	-	14.3021	14.3767	15.3192	15.6998	15.9008
Standard Error Reductions with Additional Polynomial Instruments						
Linear Model						
Intercept	-	17.69%	27.30%	28.43%	34.09%	31.88%
Log Price	-	19.40%	26.67%	28.69%	35.47%	36.10%
Log CPrice1	-	18.67%	22.01%	24.07%	35.46%	37.79%
Log CPrice2	-	13.95%	14.50%	15.62%	16.35%	25.76%
Log Dist	-	19.38%	29.08%	29.89%	35.50%	32.91%
Feature	-	34.96%	38.21%	39.02%	54.07%	54.07%
Display	-	10.80%	10.50%	10.20%	13.00%	14.70%
Non-Linear Model						
Intercept	-	19.40%	34.11%	39.56%	44.99%	47.97%
Price	-	48.17%	59.23%	58.43%	63.62%	65.85%
CPrice1	-	41.58%	57.36%	56.87%	59.27%	61.08%
Dist	-	29.32%	46.60%	49.79%	49.58%	49.06%
Feature	-	35.80%	39.49%	36.93%	36.93%	36.08%

Table 3 Efficiency Gains with Different Wavelet Filters, Decomposition Levels, Number of Instruments, Number of Exogenous Variables and Sample Sizes.

				β_0			β_1			β_2		
	Sample	*Exogenous*	*Instruments*	*Asymptotic*	*Actual*	*Estimated*	*Asymptotic*	*Actual*	*Estimated*	*Asymptotic*	*Actual*	*Estimated*
Variance of Estimates and Estimates of Variances as Proportions of Least Squares Variances (standard errors in parenthesis)												
Least Squares	*32*	*1*	*-*	*1.000*	*1.045 (1.183)*	*0.842 (0.208)*	*1.000*	*1.035 (1.088)*	*0.803 (0.221)*	*-*	*-*	*-*
D.L. Asym. - Level 0	*32*	*1*	*4*	*0.585 (0.208)*	*0.830 (1.076)*	*0.547 (0.210)*	*0.552 (0.213)*	*1.058 (1.225)*	*0.487 (0.212)*	*-*	*-*	*-*
D.L. Asym. - Level 1	*32*	*1*	*4*	*0.679 (0.193)*	*0.946 (1.146)*	*0.620 (0.203)*	*0.623 (0.205)*	*1.072 (1.188)*	*0.535 (0.209)*	*-*	*-*	*-*
D.L. Asym. - Level 2	*32*	*1*	*4*	*0.670 (0.209)*	*0.802 (0.983)*	*0.610 (0.216)*	*0.622 (0.213)*	*1.030 (1.180)*	*0.537 (0.218)*	*-*	*-*	*-*
D.L. Asym. - Level 3	*32*	*1*	*4*	*0.584 (0.209)*	*0.885 (1.135)*	*0.546 (0.203)*	*0.551 (0.218)*	*1.045 (1,221)*	*0.480 (0.206)*	*-*	*-*	*-*
D. E. Phase - Level 0	*32*	*1*	*4*	*0.455 (0.212)*	*0.714 (0.925)*	*0.434 (0.200)*	*0.483 (0.223)*	*0.999 (1.154)*	*0.424 (0.205)*	*-*	*-*	*-*
D. E. Phase - Level 1	*32*	*1*	*4*	*0.596 (0.207)*	*0.886 (1.076)*	*0.552 (0.207)*	*0.566 (0.204)*	*1.080 (1.226)*	*0.490 (0.204)*	*-*	*-*	*-*
D. E. Phase - Level 2	*32*	*1*	*4*	*0.635 (0.209)*	*0.873 (1.086)*	*0.585 (0.210)*	*0.609 (0.212)*	*1.058 (1.210)*	*0.525 (0.209)*	*-*	*-*	*-*
D. E. Phase - Level 3	*32*	*1*	*4*	*0.578 (0.224)*	*0.811 (0.979)*	*0.543 (0.220)*	*0.585 (0.228)*	*0.990 (1.103)*	*0.516 (0.225)*	*-*	*-*	*-*
Haar - Level 0	*32*	*1*	*4*	*0.348 (0.169)*	*0.668 (0.884)*	*0.346 (0.163)*	*0.330 (0.179)*	*0.954 (1.139)*	*0.306 (0.164)*	*-*	*-*	*-*
Haar - Level 1	*32*	*1*	*4*	*0.312 (0.169)*	*0.698 (0.997)*	*0.320 (0.157)*	*0.286 (0.164)*	*0.950 (1.135)*	*0.271 (0.152)*	*-*	*-*	*-*
Haar - Level 2	*32*	*1*	*4*	*0.431 (0.187)*	*0.727 (0.998)*	*0.422 (0.181)*	*0.423 (0.189)*	*0.982 (1.167)*	*0.385 (0.180)*	*-*	*-*	*-*
Haar - Level 3	*32*	*1*	*4*	*0.584 (0.212)*	*0.818 (1.119)*	*0.544 (0.207)*	*0.594 (0.221)*	*0.977 (1.198)*	*0.517 (0.216)*	*-*	*-*	*-*
Aitken	*32*	*1*	*-*	*0.012*	*-*	*-*	*0.023*	*-*	*-*	*-*	*-*	*-*
D.L. Asym. - Level 0	*32*	*1*	*10*	*0.277 (0.147)*	*0.636 (0.963)*	*0.280 (0.137)*	*0.271 (0.154)*	*0.941 (1.113)*	*0.246 (0.140)*	*-*	*-*	*-*
D.L. Asym. - Level 1	*32*	*1*	*10*	*0.327 (0.159)*	*0.687 (0.901)*	*0.335 (0.155)*	*0.298 (0.158)*	*0.942 (1.130)*	*0.280 (0.154)*	*-*	*-*	*-*
D.L. Asym. - Level 0	*32*	*1*	*15*	*0.131 (0.087)*	*0.498 (0.739)*	*0.158 (0.092)*	*0.132 (0.103)*	*0.847 (1.031)*	*0.134 (0.101)*	*-*	*-*	*-*
D.L. Asym. - Level 1	*32*	*1*	*15*	*0.183 (0.107)*	*0.669 (0.888)*	*0.207 (0.110)*	*0.175 (0.120)*	*0.913 (1.070)*	*0.173 (0.116)*	*-*	*-*	*-*
Least Squares	*32*	*2*	*10*	*1.000*	*1.048 (1.096)*	*0.536 (0.276)*	*1.000*	*1.012 (0.842)*	*0.513 (0.604)*	*1.000*	*0.987 (0.815)*	*0.580 (0.361)*
D.L. Asym. - Level 0	*32*	*2*	*10*	*0.261 (0.162)*	*0.571 (0.717)*	*0.181 (0.107)*	*0.503 (0.226)*	*0.887 (0.560)*	*0.254 (0.343)*	*0.247 (0.189)*	*0.892 (0.939)*	*0.213 (0.196)*
D.L. Asym. - Level 1	*32*	*2*	*10*	*0.216 (0.160)*	*0.438 (0.606)*	*0.195 (0.134)*	*0.488 (0.219)*	*0.922 (0.585)*	*0.259 (0.355)*	*0.290 (0.318)*	*0.846 (1.174)*	*0.270 (0.299)*
Aitken	*32*	*2*	*10*	*0.010*	*-*	*-*	*0.073*	*-*	*-*	*0.033*	*-*	*-*
Ratios of Average Actual and Estimated Variances to Asymptotic (standard errors in parenthesis)												
Least Squares	*32*	*1*	*-*	*-*	*-*	*0.842 (0.208)*	*-*	*-*	*0.803 (0.221)*	*-*	*-*	*-*
D.L. Asym. - Level 0	*32*	*1*	*4*	*-*	*1.459 (1.804)*	*0.967 (0.258)*	*-*	*2.013 (2.326)*	*0.939 (0.387)*	*-*	*-*	*-*
D.L. Asym. - Level 1	*32*	*1*	*4*	*-*	*1.410 (1.682)*	*0.955 (0.306)*	*-*	*1.765 (1.941)*	*0.919 (0.365)*	*-*	*-*	*-*
D.L. Asym. - Level 2	*32*	*1*	*4*	*-*	*1.208 (1.430)*	*0.974 (0.393)*	*-*	*1.718 (1.960)*	*0.940 (0.450)*	*-*	*-*	*-*
D.L. Asym. - Level 3	*32*	*1*	*4*	*-*	*1.564 (1.959)*	*0.993 (0.347)*	*-*	*1.997 (2.406)*	*0.945 (0.389)*	*-*	*-*	*-*
Least Squares	*128*	*1*	*-*	*-*	*-*	*0.578 (0.421)*	*-*	*-*	*0.582 (0.337)*	*-*	*-*	*-*
D.L. Asym. - Level 0	*128*	*1*	*4*	*-*	*0.847 (0.760)*	*0.485 (0.363)*	*-*	*0.960 (0.910)*	*0.462 (0.276)*	*-*	*-*	*-*
D.L. Asym. - Level 1	*128*	*1*	*4*	*-*	*0.888 (0.822)*	*0.446 (0.345)*	*-*	*0.991 (0.944)*	*0.431 (0.273)*	*-*	*-*	*-*
D.L. Asym. - Level 2	*128*	*1*	*4*	*-*	*0.841 (0.706)*	*0.437 (0.314)*	*-*	*0.960 (0.917)*	*0.435 (0.256)*	*-*	*-*	*-*
D.L. Asym. - Level 3	*128*	*1*	*4*	*-*	*0.936 (0.755)*	*0.466 (0.351)*	*-*	*1.025 (0.846)*	*0.454 (0.272)*	*-*	*-*	*-*
Least Squares	*512*	*1*	*-*	*-*	*-*	*0.541 (0.317)*	*-*	*-*	*0.548 (0.306)*	*-*	*-*	*-*
D.L. Asym. - Level 0	*512*	*1*	*4*	*-*	*0.840 (0.741)*	*0.457 (0.312)*	*-*	*0.898 (0.815)*	*0.459 (0.294)*	*-*	*-*	*-*
D.L. Asym. - Level 1	*512*	*1*	*4*	*-*	*0.723 (0.654)*	*0.355 (0.235)*	*-*	*0.767 (0.736)*	*0.365 (0.225)*	*-*	*-*	*-*
D.L. Asym. - Level 2	*512*	*1*	*4*	*-*	*0.744 (0.732)*	*0.376 (0.252)*	*-*	*0.810 (0.770)*	*0.389 (0.244)*	*-*	*-*	*-*
D.L. Asym. - Level 3	*512*	*1*	*4*	*-*	*0.768 (0.739)*	*0.411 (0.293)*	*-*	*0.858 (0.843)*	*0.421 (0.276)*	*-*	*-*	*-*

References

[1] Abramovich, F., Bailey, T.C. & Sapatinas T. (2000). Wavelet analysis and its statistical applications, *The Statistician*, **49**, 1–29.

[2] Antoniadis, A., Bigot, J. & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, **6**, Issue 6, 1–83.

[3] Bronnenberg, B.J., Mela, C.F. & Boulding, W. (2006). The periodicity of pricing. *Journal of Marketing Research*, **43**, 477–493.

[4] Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34**, 305–334.

[5] Cragg, J. (1983). More efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* **51**, 751–763.

[6] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

[7] Davidson, R. & MacKinnon, J.G. (2004). *Econometric Theory and Methods*. New York: Oxford University Press.

[8] Dagenais, M.G. & Dagenais, L.D. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, **76**, 193–221.

[9] Donald, S.G., Imbens, G. & Newey, W. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, **117**, 55–93.

[10] Fan, Y. & Gencay, R. (2007). Unit root and cointegration tests with wavelets. *Preprint*, Simon Fraser University.

[11] Gencay, R., Selcuk, F. & Whitcher, B. (2001a). Scaling properties of foreign exchange volatility. *Physica A*, **289**, 249–266.

[12] Gencay, R., Selcuk, F. & Whitcher, B. (2001b). Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A*, **289**, 543–556.

[13] Gencay, R., Selcuk, F. & Whitcher, B. (2001c). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, San Diego: Academic Press.

[14] Gencay, R., Selcuk, F. & Whitcher, B. (2003). Systematic risk and time scales. *Quantitative Finance*, **3**, 108–116.

[15] Gencay, R., Selcuk, F. & Whitcher, B. (2005). Multiscale systematic risk. *Journal of International Money and Finance*, **24**, 55–70.

[16] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.

[17] Hellerstein, J.K. & Imbens, G.W. (1999). Imposing moment restrictions from auxiliary data by weighting. *The Review of Economics and Statistics*, **81**, 1–14.

[18] Hong, Y. & Lee, J. (2001a). Testing serial correlation of unknown form via wavelet methods. *Econometric Theory*, **17**, 386–423.

[19] Hong, Y. & Lee, J. (2001b). One-sided testing for ARCH effects using wavelets. *Econometric Theory*, **17**, 1051–1081.

[20] Imbens, G. (1997). One-step estimators for over-identified generalized method of moment models. *Review of Economic Studies*, **64**, 359–383.

[21] Imbens, G.W., Spady, R.H. & Johnsons, P. (1998). Information theoretic alternatives to inference in moment condition models, *Econometrica*, **66**, 333–357.

[22] Johnson, R.A. & Wichern, D.W. (1998). *Applied Multivariate Statistial Analysis*. 4th edition, New Jersey: Prentice Hall.

[23] Leeflang, P.S. & Witting, D.R. (1992). Diagnosing competitive reactions using (aggregated) scanner data. *International Journal of Research in Marketing*, **9**, 39–57.

[24] Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional aata are available, with an application to patents and R&D. *Econometrica*, **65**, 1201–1213.

[25] Mallat, S.G. (1989), A theory for multiresolution signal decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

[26] Nason, G.P. & von Sachs, R. (1999). Wavelets in time-series analysis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **357**, 2511–2526.

[27] Nason, G.P. & Sapatinas, T. (2002). Wavelet packet transfer function modelling of nonstationary time series. *Statistics and Computing*, **12**, 45–56.

[28] Nevo, A. (2002). Sample selection and information-theoretic alternatives to GMM. *Journal of Econometrics*, **107**, 149–157.

[29] Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *Journal of Business and Economic Statistics*, **21**, 43–52.

[30] Newey, W.K. & West, K.D. (1987). A simple, positive and semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

[31] Peracchi, F. (2001). *Econometrics*. Chichester: John Wiley & Sons.

[32] Percival, D.B. & Walden, T.W. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

[33] Priestley, M.B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

[34] Qian, H. & Schmidt, P. (1999). Improved instrumental variables and generalized method of moments estimators. *Journal of Econometrics*, **91**, 145–169.

[35] Ramsey, J.B. & Lampart, C. (1998a). The decomposition of economic relationships by time scale using wavelets: money and income. *Macroeconomic Dynamics*, **2**, 49–71.

[36] Ramsey, J.B. & Lampart, C. (1998b). The decomposition of economic relationships by time scale using wavelets: expenditure and income. *Studies in Nonlinear Dynamics and Econometrics*, **3**, 23–42.

[37] Ramsey, J.B. & Zhang, Z. (1997): The analysis of foreign exchange rates using waveform dictionaries. *Journal of Empirical Finance*, **4**, 341–372.

[38] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.

[39] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

[40] Woodridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.