Money demand equation continues to attract attention of econometricians with a new wrinkle provided by cointegration. We use projection pursuit (PP) regressions pioneered by Friedman and Stuetzle (1981) to suggest new estimates of partials of conditional expectations of the regressands with respect to the regressors and prove their consistency. Since the usual cointegration methodology involves linear relations, carefully chosen directions where generalized additive structure is preserved by PP methods is more flexible. These methods are computationally demanding, since the bootstrap needs to be used for confidence statements. Our numerical estimates with 18 regressors yield narrow confidence intervals.

KEY WORDS: Nonnormality, Macroeconomics, Bootstrap, Discrimination, Curse of Dimensionality.

1. INTRODUCTION

This paper is concerned with nonparametric regression and cointegration for modeling economic equilibria. In macroeconomics, cointegration has come to mean that nonstationary integrated variables of order d, denoted by I(d), have a linear equilibrium relation which follows a stationary process of order I(0). Although the linearity assumption may be problematic, attempts to introduce nonlinear nonparametric regressions are rare. This is one such attempt. We estimate Saikkonen (1991) cointegrating regression equation with a PP nonparametric method.

Economists use the log, Box-Cox transformations and its extensions (e.g., Wooldridge, 1992) to estimate nonlinear relations between a nonnegative variable y and a set of regressors x. The focus of attention is on the conditional mean E(y|x) and its derivatives. We view the projection pursuit (PP) method by Friedman and Stuetzle (1981) as a computer intensive extension of the Box-Cox model. The advantages of the PP method over other types of nonparametric regressions include: (i) PP is less sensitive to the “curse of dimensionality” and (ii) PP permits interactions compared to some generalized additive models, Hastie and Tibshirani (1987). The disadvantages are: (i) the computational burden is high, (ii) PP may not work for all geometrical shapes according to Donoho and Johnstone (1989), and (iii) PP does not provide convenient estimates of regression coefficients (elasticities) and their standard errors. This paper reports estimates of elasticities of the US long run money demand equation including their standard errors from PP regressions. Section 2 briefly reviews cointegration and PP methods. Section 3 applies PP to a money demand equation and reports results for US data including confidence intervals based on a bootstrap.

* E-mail: VINOD@murray.FORDHAM.edu. I thank James Stock for valuable help.

2. A BRIEF REVIEW OF COINTEGRATION AND PP METHODS

Although cointegration literature is vast and growing, our brief review uses the bivariate case of two time series $y_{1t}$ and $y_{2t}$ with T observations related by the system of equations:

$$y_{1t} = 3y_{2t} + u_{1t}, \quad \text{where } u_{1t} \sim \text{WN}(0, 5^2)$$  \hspace{1cm} (2.1)

$$y_{2t} = y_{2t-1} + u_{2t}, \quad \text{where } u_{2t} \sim \text{WN}(0, 5^2)$$  \hspace{1cm} (2.2)

where WN denotes white noise stochastic process and $u_{1t}$ and $u_{2t}$ are uncorrelated. Such error term assumptions are common in the cointegration literature. Note that $y_{2t}$ is the first difference operator, implies that $y_{2t}$ is a random walk. From (2.1), note that

$$y_{1t} = 3y_{2t} + u_{1t} = [3u_{2t} + u_{1t}] - u_{1,t-1} - u_{1,t-1}$$  \hspace{1cm} (2.3)

Thus both $y_{1t}$ and $y_{2t}$ are nonstationary integrated processes of order 1, denoted by I(1). In the absence of cointegration, a linear combination like $y_{1t} - 3y_{2t}$ is also I(1). In Econometrics, two series are often...
dynamically linked due to economic equilibrating (market) forces unleashed by economic agents, such as arbitrage. For example, two time series of prices at two locations (gold prices in London and New York) arbitrage by gold traders ensures that they cannot drift too far apart from each other. If their linear combination (price difference) is stationary, or \( I(0) \), they are said to be cointegrated and we write:

\[
z_t = y_{1t} - 3y_{2t} - \mu \quad I(0) \quad (2.4)
\]

In matrix notation, \( y_t = (y_{1t}, y_{2t}) \) is cointegrated with the cointegrating vector \( a^\top = (1, -3) \). Now, if \( z_t = a^\top y_t \), \( \mu \) is stationary and ergodic in second moments, it can be shown, Hamilton (1994, p. 587) that a (super) consistent estimate of \( a^\top \) can be found by minimizing \( T^{-1} \sum_{t=1}^{T} z_t^2 \), i.e., by ordinary least squares (OLS). The OLS estimate is consistent even when the errors are serially correlated. A multivariate generalization assuming exactly one cointegrating relation is called “triangular representation.” Let \( \mathbf{z} \) and \( y_{2t} \) and \( y_{1t} \) denote \( g \) \( \mathbb{C} \) 1 vectors and \( \mathbf{3} \) and \( \mathbf{3}_0 \) denote an intercept. A \( g \)-variate system is:

\[
y_{1t} = \mathbf{3}_0 + \mathbf{3}^\top y_{2t} + z_t^* \quad \text{?} \quad y_{2t} = u_{2t}, \quad (2.5)
\]

where the errors are assumed to satisfy:

\[
(z_t^*, u_{2t}) \sim <(L) \% (2.6)
\]

where \( L \) denotes the lag operator, \( <(L) \) is an \( n \times 1 \) vector with \( n = g + 1 \), \( % \) is an \( n \times 1 \) independent and identically distributed (iid) vector with zero mean, finite 4-th moments and positive definite covariance matrix \( E(%) \). Also assume that the sequence of matrices \( \{s<\} \) for \( s = 0 \) to \( \infty \) is absolutely summable and the rows of \( <(1) \) are linearly independent. Then the OLS estimates \( \mathbf{3}_0 \) and \( \mathbf{3} \) are consistent and converge in law to an expression involving \( n \)-dimensional Brownian motion, Hamilton (1994, p. 588).

In general, the errors in (2.6) and OLS estimates have a nonstandard nonnormal distribution because of the “nuisance” correlations between \( z_t^* \) and \( u_{2t} \). If we are willing to assume that the correlation between \( z_t^* \) and \( u_{2t-s} \) is zero for \( |s| > w \), then including \( w \) leads and \( w \) lags in (2.5) solves the problem, Saikkonen (1991).

\[
y_{1t} = \mathbf{3}_0 + \mathbf{3}^\top y_{2t} + D_{s=-w}^w (!) \mathbb{P}_{y_{2t-s}} + z_t^* \quad (2.7)
\]

where OLS estimates of \( \mathbf{3}_0 \), \( ! \), \( \mathbf{3}^\top \) and errors \( z_t^* \) are well behaved. If \( y_{2t} \) vector contains \( k \) regressors, there are \( k(2w+1) \) additional regressors \( y_{2t-s} \). Since dynamic economic relations are linear only under restrictive assumptions, one can permit nonlinearities in (2.7) by using nonparametric regression surveyed in Ullah and Vinod (1993). They refer to over a dozen applications from labor economics to hazardous waste. Since (2.7) must have several regressors, it is well known that kernel-type regressions will yield unreliable signs and magnitudes of coefficients, by the “curse of dimensionality”. This paper shows that the PP regressions (PPRs) can provide meaningful parameter estimates.

Multivariate data are often simplified by considering lower dimensional “projections,” such as Fisher’s linear discriminant analysis. Fisher used the ‘classification error rate’ criterion and analytically minimized it by assuming normally distributed samples with equal covariance matrices. The PP avoids Fisher’s assumptions and uses numerical solutions. First, one computes linear ‘projections’ as smoothing operations viewed as shadows of structures in their full dimensionality. Second, the ‘pursuit’ part numerically maximizes the amount of structure present in that projection. Friedman and Stuetzle’s (1981) PP approximation for nonparametric regression is the main focus here. We use subscripts \( i = (i_1, \ldots, i_M) \) for projection (or pass) numbers, \( j = (j_1, \ldots, j_k) \) for regressors and often suppress the subscript \( t = (t_1, \ldots, T) \) for observations. Denote the dependent variable by \( y \), the \( j \)-th regressor by \( x_j \), the error by \( % \) and the estimators by hats(\( ^\)\( )\)). Let \( W_i = D_{j=1}^k "_{ij} x_j = "^i_k \) represent the \( i \)-th linear projection, where \( "_i = ("_i1, "_i2, \ldots, "_ik \) is of unit length, \( "^i_k \) is built from \( x_{ji} \). By contrast, in the standard notation \( E(y|X=x) \), the \( X=(x_{1j}, x_{2j}, \ldots, x_{kj}) \) is built from \( x_{ji} \). Let \( \# \) denote a \( M \times 1 \) vector containing the scaling parameters \( \# \) and \( s_i(W_i) \) denote numerically obtained generally nonlinear “smooth functions”. The PPR equation is

\[
y = F(y, W_i, \#) = y + D_{i=1}^M \# s_i(W_i) + \% \quad E(\%)=0 \quad (2.8)
\]
The \( y = \mathbb{E}(y) \) is estimated by the sample mean \( y = T^{-1} \sum_{t=1}^{T} y_t \). \( \mathbb{E}(y|x) \) is \( y + D_{i=1}^{M} \hat{s}_i(W_i) \). For \( i \)-th projection, the smooth functions \( s_i \) are standardized to have zero mean and variance unity:

\[
\mathbb{E}(s_i(W_i)) = 0 \quad \text{and} \quad \mathbb{E}(s_i^2(W_i)) = 1 \quad \text{for all} \quad i = 1, 2, \ldots, M (2.9)
\]

The smoothing in PPRs generalizes the Box and Cox (1964) transformation defined as:

\[
y_t^\gamma = \begin{cases} y_t & \text{if } \gamma = 1; \\
y_t^{1/\gamma} & \text{if } \gamma > 0; \\
\log(y_t) & \text{if } \gamma = 0 \end{cases} \quad (2.10)
\]

which in turn generalizes the log transformation. In a nonparametric PP setting, we do not try to coax the nonlinear relation into any analytically tractable or known mathematical form. PPR simply smooth the scatterplot, nonparametrically tracing a smooth curve, whatever its shape (e.g., Fig. 5). The smoothing algorithm makes \( M \) passes over running medians, estimates variability at each point and smooths it by a fixed width moving average. It further smooths the running medians by “locally linear” fits and bandwidths determined by the smoothed local variance. The usual econometric approach cannot delineate nonlinearities similar to Figures 5 and 6, since they lack analytical expressions.

Unlike traditional smoothing, the PPR algorithm uses cross-validation and has the advantage that, in general, it uses \( M > 1 \) passes over the data with \( M \) functions \( s_i(W_i) \) along \( M \) directions. Hence it can eventually capture all of the relevant structure. Note that orthogonal projections imply zero covariances and help define ‘covariance estimators’. Hence Duan (1990, eq. 1.7) defines PPR by:

\[
\text{Cov}[x, y - \mathbb{E}(y|W)] = 0 \quad (2.11)
\]

In practice, the algorithm's first pass computes a linear regression of \( (y - \hat{y}) \) on a function \( s_1(W_1) \) and computes the residuals \( r_{1t} \) ordered by ascending values of \( W_1 = x \). It finds \( \hat{y} \) by maximizing the ‘explained’ proportion of the variance of \( y \) decomposing \( y \) into the ‘smooth’ \( s_1(W_1) \) and the residual \( r_1 = (r_{1i}) \). The second pass regresses \( r_1 \) on \( s_2(W_2) \), the \( i \)-th pass regresses \( r_{i-1} \) on \( s_i(W_i) \), terminating when all the relevant structure is captured. The \( i \) determines the \( M \) in (2.8) when the fit is good enough for a user specified threshold. The PPR is obviously computer intensive, since it numerically searches over all possible unit directions, and eventually minimizes (over all possible \( #, s_i \) and \( "_i \)) the mean squared error (MSE):

\[
\mathbb{E}[(y - \hat{y})^2] = 0.5 \sum_{i=1}^{M} \mathbb{E}[(y - \hat{y})^2] (2.12)
\]

Diaconis and Shahshahani (1984) prove that for large enough \( M \), PPR can approximate (uniformly on compact sets) arbitrary continuous functions of the regressors. To consider the likelihood function for (2.12) assume spherically symmetric standard normal errors at each pass. Now insert the subscript \( t = 1, \ldots, T \) for \( y, x_j \) and \( W_i \). Then \( W_i = D_{j=1}^{k} "_i x_{jt} \) and the corresponding sample log likelihood function is:

\[
L_i = \text{constant} \sum_{t=1}^{T} \mathbb{E}[y_t - \hat{y} | x_j] (2.13)
\]

The \( k \) \( \mathcal{C}_1 \) score vector of the \( i \)-th projection (\( i = 1, \ldots, M \)) is:

\[
\text{score}(x) = -L_i/\sqrt{x_j} (2.14)
\]

Note that the score function will depend on \( s_i(W_i) \) and \( "_i = W_i/\sqrt{x_{jt}} \), of which the smooth function \( s_i \) may have no known functional form. Hall (1989) and Zhu and Fang (1992) use a derivation involving Kernel functions to express \( s_i \) in their proofs of consistency and their derivations of rates of convergence. For simplicity, let us assume the Taylor Theorem regularity conditions, Gallant (1987, p. 13). We also assume that \( s_i(W_i) \) can be approximated by the linear expression \( [k_i + W_i] \), where the constant \( k_i \) changes with \( t \). Now the average derivative is

\[
\mathbb{E}[(s_i(D_{j=1}^{k} "_i x_{jt})/\sqrt{x_{jt}} \prod "_i (2.15)]
\]
Next, we turn to derivation of two Theorems regarding estimating partials of \( E(y|x) \) from the estimates of \( \# \) and "\( ^{\text{w}}E(\text{ji}) \). First, we restate Stoker's (1986) assumptions and his condition A in our notation.

**Assumption 1**: \( H \) is a convex subset of \( d^M \) with nonempty interior. The underlying measure \( \mu \) can be written in product form as \( \mu = \mu_1 \mu_2 \cdots \mu_x \), where \( \mu_x \) is Lebesgue measure on \( d^M \) which implies that components of \( x \) are not perfectly correlated.

**Assumption 2**: \( s_j(x) \) is continuously differentiable in the components of \( x \) for all \( x \) in the interior of \( H \) and \( E[\text{score}_r(x)] \) and \( E[\text{score}_r(x) \text{score}_r(x)] \) exist for all \( i=1, 2, \ldots, M \).

**Assumption 3**: For \( x \in \partial H \), where \( \partial H \) is the boundary of \( H \), we have \( s_i(x)=0 \).

**Condition A**: \( G(x) \) is continuously differentiable for all \( x \in \partial H \), where \( \partial H \) differs from \( H \) by a set of \( \mu \)-measure 0. \( E(y) \), \( (G/x) \) and \( E(\text{score}(x) y) \) exist for \( i=1, \ldots, M \) and \( j=1, \ldots, k \).

**Assumption 4**: (a) \( (y, F) \) of (2.8) satisfy condition A. \( E(F/W) \) are nonzero for \( i=1, \ldots, M \).
(b) \( (x, x) \) satisfies condition A for each \( j=1, \ldots, k \).

Zhu and Fang (1992) relax Hall's assumptions and prove global consistency and rates of convergence \( q/2 \). They still need spherically symmetric \( x \) and non-atomic distributions. The following additional assumption is needed only for cointegration applications.

**Assumption 5**: Zhu and Fang's (1992) results continue to apply for linear cointegrating models, which are preserved under linear projections and linear augmentation of I(0) variables for nonlinear models.

**Theorem 1**: The PPR above gives \( \hat{\#}^{\text{w}} \) as a consistent estimator of "\( ^{\text{w}}\# \) when " is \( M \) \( \subset \) \( k \) and \( \# \) is \( M \) \( \subset \) 1. Proof: The consistency of the PPR estimates \( \hat{\#} \) and \( \hat{\#} \) under appropriate conditions has been shown by many, including Hall (1989). Hence, by Slutsky's theorem, we have consistency of \( \hat{\#} \) \( \# \).

**Theorem 2**: Given assumptions 1 to 4 and condition A, the partials of \( E(y|x) \) estimated by \( \hat{\#} \) satisfy:
\[
E[\hat{\#}(\text{score}(x))]/x \rightarrow E[\#]/x \quad (\#) (2.16)
\]
Proof: See Stoker's theorem 1. He uses integration by parts to establish a link between partials of \( E(y|x) \) and covariance estimators like (2.11). Stoker's single index models have "instrumental variables estimators" obtained by using \( \text{score}(x) \) as instruments. Letting \( i=1, \ldots, M \) yields an M-index model. From the chain rule, (2.8) and (2.15), the following must be satisfied for each \( t: \) \( E(y|x)/x = D^M_{i=1} \#(\text{score}(x))/xj \) \( \# \) \( \longleftarrow \) \( ^{\text{w}}\# \) \( \longleftarrow \) \( ^{\text{w}}\# \). Derive (2.16) by combining them over all \( j \) and \( t \).

Vinod and Ullah (1988) argue that the focus of attention in econometric regressions is usually on what they call "amorphous" partial derivatives (apd). Their notation apd\( (y, x) = E(y|x)/x \) is convenient. Wooldridge (1992) also argues that the conditional mean \( E(y|x) \) is the function of interest when economists use "behavioral" partial derivatives. Since the PP methods generalize the Box-Cox transformation, Wooldridge's arguments in favor of partials of \( E(y|x) \) apply here. We first make a logarithmic transformation in the definition of \( y \) and \( x \) to be able to interpret the partials as elasticities.

Having shown consistency and defined our apd's, the next step is statistical inference regarding the apd's and confidence intervals using 1000 bootstrap resampled apd estimates. In (2.7) there are \( k(2w+1) \) extra regressors \( y_{2t-4} \). For our example, we have \( k=3 \) and \( w=2 \) (two leads and lags) leading to 15 extra regressors. For kernel methods the curse of dimensionality would have been impossible to overcome here. The PPR algorithm in S-PLUS (1993) software needed only about twelve hours on a (slow) 75 mhz Pentium computer to resample our three elasticities 1000 times. Bootstrap refinements to allow for dependent data are given in Vinod (1996). Since our inference in the money demand cointegration largely agrees with the literature, we can also avoid refined bootstrap inference using so-called estimating functions from Vinod (1998).

Using the PPR to estimate partials of \( E(y|x) \) and using the bootstrap for inference are new. The PPR involves projecting the regressor matrix in \( M \) carefully chosen directions where generalized additive
structure is preserved. A practical advantage for estimation of behavioral partials from (2.16) is that the point at which the partial is measured does not need to be specified. The additive linearity on PPRs and the approximation (2.15) help avoid this need. On the other hand, if one wishes to specify the point \( x_0 \), the formulas (2.15) and (2.16) will lose some of their simplicity. By contrast, in the usual nonlinear regression, where the mathematical functional forms are specified, one needs to be explicit about the point where to evaluate the partials. Computer intensive PPR with the bootstrap based inference has become feasible due to declining computer costs. Even further bootstrap refinements, Vinod (1993, 95, 96, 97, 98), are feasible, which was not true in 1981 when the PP was first proposed.

3. MONEY DEMAND EQUATION FOR THE US DATA

Monetary policy requires reliable estimates of the long-run demand for money, Judd and Scadding (1982). Two controversial issues in this literature are: (i) Is the long-run demand equation stable? (ii) What are the magnitudes of related elasticities? This section illustrates our PPR methods with Stock and Watson's (1993) example to estimate a (cointegrating relation) a money demand equation. We use their annual data for 1900 to 1989 with \( T=90 \). The (natural) log of M1 money supply is denoted by \( m \), log of net national product (NNP) is denoted by \( y \), log of price deflator for NNP is denoted by \( p \), and the rate on 6-month commercial paper is denoted by \( r \).

Our money demand equation is based on (2.7), where \( w=2 \) leads and lags of first differences are included, designed to allow for the nuisance correlation problem. Hence, we cannot use the usual t tests on (2.7) to eliminate the lead and lag terms. Our nonlinear nonparametric regression of \( m \) on \( p \), \( y \) and \( r \), along with two leads, two lags and current level of the first differences of the three regressors has eighteen regressors. Note that we are mainly interested in the first three regression coefficients interpreted as partial derivatives of the dependent variable \( m \) with respect to the three regressors \( p \), \( y \) and \( r \). However, we are facing the curse of dimensionality with 18 regressors.

The user of PPR software needs to choose a range for \( M \), the number of directions. We choose \( M_1 \) with \( M=10 \), since it yields a better-behaved model and residuals. Denote by rss the ‘residual sum of squares’ and by ssrac9 the ‘sum of squares of residual autocorrelations’ of orders 1 to 9. As \( M_1 \) increases, the rss values decline monotonically as the number of parameters increases. For \( M_1=9,10,11,12 \) we have rss= (0.0006763, 0.0006139, 0.0004223, 0.0002635) and ssrac9= (0.2104, 0.0350, 0.0587 and 0.1997), respectively. Thus, ssrac9 has a local minimum at \( M_1=10 \). It is convenient to report the corresponding elasticities after the notation in terms of \( \gamma \)’s is established. The low values of rss suggest a very good fit compared to typical least squares values reported in the literature and may mean less severe post-1973 overpredictions noted by Judd and Scadding (1982). To assess the effect on the fit of the last 15 regressors, we fitted the PPR without the 15 regressors and found that the rss= 0.0003 for \( M_1=5 \) was similar to the rss=0.0002635 observed for \( M_1=12 \).

Fig. 9 displays our 90-year annual data against time. The dashed series with maximum ups and downs, which starts high on the left has a u-shaped dip in the middle and again goes high toward the end represents the interest \( r \). The solid line is for \( m \), the dotted line is for \( p \) and the one with dots and dashes is for \( y \). Fig. 1 has \( p \), Fig. 2 has \( y \) and Fig. 3 has \( r \) along the horizontal axis and all three scatter plots have \( m \) on the vertical axis. We plot “the smooth” of \( m \) against the first projection (weighted average) of the 18 regressors in Fig. 4. It uses the Cov[\( x, r - E(r| x) \)]=0 defined by (2.11). Fig. 5 plots the obviously nonlinear smooth of \( m \) against the second best weighted average of the 18 regressors using Cov[\( x, r - E(r| x) \)]=0, where \( r_1 \) is the first pass residual. It attempts to maximize the proportion of the variance of \( r_1 \) explained by the nonparametric smooth \( s_2 \). Similarly, Fig. 6 plots the smooth of \( m \) against the third best linear combination, using Cov[\( x, r - E(r| x) \)]=0. Although the “principal components regression” also seeks best weighted averages, the principal axes must be orthogonal to each other and one uses a regressor covariance matrix \( \hat{A} \) not smoothing of the scatter of residuals. Clearly, the PPR is flexible,
with less restrictive assumptions and uses cross validation of smoothed scatter plots similar to some nonparametric regressions. To assess the quality of our fit, Fig. 7 plots \( m \) against its fitted values and shows a very good fit, confirming the result based on rss and noted above. Fig. 8 plots the final residuals against the fitted values to assess randomness. The graphical analysis highlights the nonlinearity and good performance of PPR here.

The asymptotic theory (Wald tests, etc.) of Stock and Watson's dynamic ordinary least squares (DOLS) estimator of linear cointegrating relations is sophisticated. However, our PPR methods are needed to supplement DOLS by incorporating possible nonlinearity of cointegrating relations (See Fig. 5 and Fig. 6). We also compare the elasticities from DOLS with amorphous partial derivatives from PPR. The \( \text{apd}(m, p) \) represents price elasticity, \( \text{apd}(m, y) \) represents income elasticity, and \( \text{apd}(m, r) \) represents interest semi-elasticity. For example, \( \text{apd}(m, p) \) measures the % change in money demand with respect to a 1% change in prices.

The S-PLUS-based bootstrap function developed by Efron and Tibshirani (1993) was used to replicate the PPRs 1000 times. The GAUSS computer language was also used in our analysis. For brevity, plots of histograms based on 1000 estimates of the three apd's are not reproduced. Let us denote min=minimum, Q1=first quartile, Q3= third quartile, and max=maximum from the 1000 estimates.

Our bootstrap yields: [min, Q1, Median, Mean, Q3, max] for \( p = [0.4174, 0.7317, 0.7984, 0.7940, 0.8565, 1.132] \), for \( y = [0.2766, 0.5642, 0.6266, 0.6249, 0.6882, 0.9104] \), and for \( r = [-0.09362, -0.07714, -0.07107, -0.07051, -0.06455, -0.03932] \). Also, [Skewness, Kurtosis, Standard Error (Standard Deviation)] are: \( p = [-0.260, 3.362, 0.09295] \), \( y = [-0.233, 3.058, 0.09546] \) and \( r = [0.372, 2.973, 0.09296] \) where the Skewness and Kurtosis are the usual Pearsonian measures and the standard deviation is used as the standard error for inference purposes. Although the standard error is the largest for \( y \), it is smaller than 0.12 to 0.27 reported by Stock and Watson (1993, p. 801). The PPR appears to have succeeded in getting somewhat more reliable estimates of income elasticity.

Efron and Tibshirani (1986) and Vinod's (1993) survey describe when normalized bias corrected (BC) confidence intervals are correct and discuss further refinements for dependent data. Now we report the PPR estimates bracketed by the corresponding Lower and Upper limits of the 95% normalized BC bootstrap confidence intervals. The [Lower limit, PPR, Upper limit] are \( p = [0.7788, 0.8763, 1.1323] \), \( y = [0.4983, 0.6621, 0.8500] \), and \( r = [-0.0912, -0.0742, -0.0584] \). Similar DOLS values are: \( y = [0.88, 0.97, 1.06] \) and \( r = [-0.127, -0.101, -0.075] \). Since these are computed subject to the assumption that \( p = 1 \), which may have been needed to reduce the curse of dimensionality DOLS values for interest elasticity is not available. Stock and Watson (1993, p. 801) note that \( p \) do not differ from 1 at the 10% level. In our case, 1.13, the upper limit of a bootstrap 95% interval, is larger than 1 and supports (at 5% level) this result. Our PPR estimate 0.6621 for \( y \) is subject to a smaller bootstrap standard error and shorter 95% interval. Note that a 1% increase in income leading to a 0.66% increase in M1 money demand (demand deposits at banks plus currency, traveller’s checks and other checkable deposits) is plausible, when we note that a large proportion of middle and low income households in the US carry a large debt burden. The PPR estimate of \( r \) is –0.0742, also close to the DOLS value of –0.10. The comparisons between DOLS and PPR suggest that PPR method yields reasonable estimates.

If simple linear cointegrating relations are desired, our approach provides a Taylor series approximation: \( \hat{E}(y|x) \prod \cdot y^j \cdot D_{x_j}^k \cdot \hat{j}_{x_j} \), where \( \hat{j}_{x_j} \) are PPR estimates of partials with respect to \( x_j \). Granger and Hallman (1991) discuss the limitations and difficulties in proposing cointegration tests that are invariant to monotone data transformations. Despite the analytically intractable nonlinearity of Figures 5 and 6, the problems of invariant tests can be postponed for such Taylor approximations. It is possible that the PP methods are largely ignored in econometrics because of its distinct (unlike the usual
X") notation, computational difficulties, a lack of estimates of quantities of economic interest (similar to our partials from eq. 2.16), and a lack of confidence intervals (statistical inference). We have shown how to reduce these difficulties.

Our results are consistent with the Stock-Watson result of a single (nonlinear) stable long run demand for money. More importantly, we introduce flexibility to ameliorate the limitation that the usual (including DOLS) cointegrating relations must be linear. With rapidly declining computer costs computer intensive PP methods deserve further attention, especially in light of our two new suggestions:

(a) to interpret (\hat{\alpha}^W) from PPRs as behavioral partials from (2.16), and (b) to use the bootstrap for PPR inference and confidence statements.

Unfortunately, a comprehensive comparison among the fast expanding list of cointegration estimators and bootstraps requires massive simulations. Hence, while we use cointegration to illustrate the PP methods, we cannot claim that it is the best method for (nonlinear) cointegration. We may claim, however, that the PP methods as extended here have a general applicability for many nonlinear models in econometrics. An important empirical question is: Would nonlinear nonparametric estimation reverse the conclusions based on models that are linear in parameters? The answer is negative for the money demand example discussed in the text, as well as for the labor market discrimination example not reported here.

Further econometric theory, simulations, bootstraps and additional applications of PPR are clearly feasible and desirable. We demonstrate the following results. (i) The PP methods carefully choose the best directions, where generalized additive structure is preserved during each pass. They yield a linear combination of smooth functions \hat{s} and good fits. (ii) The PPR is helpful in solving the curse of dimensionality, since it yields meaningful signs and relative magnitudes, even with eighteen regressors present in (2.7). The additional regressors are required for proper estimation of cointegration despite certain nuisance correlations among errors. (iii) When combined with the bootstrap, our specific PPR estimate of income elasticity of money demand (\hat{\ell}_y) is shown to be rather well estimated, in the sense of having a low standard error and a short 95% confidence interval. At a broader level, the data do support a stable long-run nonparametric nonlinear money demand equation.

Econometricians often need estimates of partial derivatives of E(y|x) when there are many regressors and small data sets, Wooldridge (1992). In highly nonlinear situations and moderate to large samples, we show the practical potential of our two theorems regarding the PPR. Thus, this paper supplements the results in nonparametric regression literature, for econometricians.

REFERENCES

Fig. 1 Scatterplot with m on vertical axis and p on horizontal axis.

Fig. 2 Scatterplot with m on vertical axis and y on horizontal axis.
Fig. 3 Scatterplot with m on vertical axis and r on horizontal axis.

Fig. 4 Smooth of m against the 1st best weighted average of 18 regressors.

Fig. 5 Smooth of m against 2nd best weighted average of 18 regressors.

Fig. 6 Smooth of m against 3rd best weighted average of 18 regressors.

Fig. 7 Goodness of Fit (y on vertical axis and fitted y on horizontal axis)

Fig. 8 Residual on vertical axis and fitted y on horizontal axis.

Fig. 9 Data against time: r (most unstable line), p (dotted), m (solid) and y (dots and dashes)