The University of Manchester

Economics Discussion Paper Series EDP-0921

Adaptive Instrumental Variable Estimation of Heteroskedastic Error Component Models

Eduardo Fé-Rodríguez

November 2009

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

Adaptive Instrumental Variable Estimation of Heteroskedastic Error Component Models

Eduardo Fé-Rodríguez*

University of Manchester

November 11, 2009

Abstract

This note extends the HT estimator to allow for heteroskedastic individual effects. Unlike in previous contributions we study if Local Polynomial Regression can aid to increase efficiency attainment.

Key Words: *Hausman-Taylor, Nonparametric Regression, Heteroskedasticity.*

JEL Classification: *C14, C23.*

*Correspondence to: Eduardo Fé-Rodríguez. Health Economics. Room 1.308, University Place, Block 3. University of Manchester. Manchester, M13 9PL, UK. email:e.fe-rodriguez@manchester.ac.uk.

1 Introduction

Estimation of panel data when the error components are heteroskedastic is a topic of interest in econometrics, given that the assumption of equally distributed errors is oftern unattainable in practice. Contributions to this area have been made by Mazodier and Trognon (1978), Baltagi and Griffin (1988), Li and Stengos (1994), and Roy (2002) among many others. The last two articles consider adaptive Fixed Effects (FE) and Random Effects (GLS) estimation, so that no structural assumptions are made regarding the precise nature of the heteroskedastictiy.

There are situations when the exogeneity assumption underpinning RE in unsustainable, but FE proves limited in scope because the researcher is interested in the effect of variables which do not vary as time goes by. A classic example is estimation of wage equations, where attention focuses on years of schooling, which is constant within individual observations. In these situations the linear papel data estimator proposed by Hausman and Taylor (1981) (referred to as HT in what follows) provides efficient, consistent estimates of the coefficients in the regression, under weak regularity conditions. This note extends the HT estimator to allow for heteroskedastic individual effects. This case is of particular relevance in the analysis of Stochastic Frontiers, (Aigner et al. (1977) and Schmidt and Sickles (1984)), where time invariant heterogeneity captures individual firm inefficiency and one might expect that the spread of this distribution varies widely across the sample, with groups of small firms exhibiting wider ranges of efficiency scores than larger groups of larger firms.

As in Li and Stengos (1994) and Roy (2002), the method relies on nonparametric regression in order to estimate the conditional variance of the compound error component. Unlike these references we explore further the potential efficiency gains that could result from alternative nonparametric regression techniques. In particular, we study if Local Polynomial Regression can aid to increase efficiency attainment.

2 Adaptive IV Estimation

The point of departure of the analysis that will follow is the linear panel data estimator proposed by Hausman and Taylor (1981). These authors consider a linear panel data model such as

$$y_{it} = x'_{1,it}\beta_1 + x'_{2,it}\beta_2 + z'_{1,i}\alpha_1 + z'_{2,i}\alpha_2 + u_i + v_{it}$$

for $i = 1, \ldots, N$, $t = 1, \ldots, T$, which may be written in matrix form as

$$\begin{aligned} y &= X_1\beta_1 + X_2\beta_2 + Z_1\alpha_1 + Z_2\alpha_2 + u + v. && (2.1)\\ &= Q\gamma + \varepsilon && (2.2) \end{aligned}$$

where $\varepsilon = u + v$. In this model, $x_{1,it}$ and $x_{2,it}$ are K_1 and K_2 vectors of time varying regressors, $z_{1,i}$ and $z_{2,i}$ are L_1 and L_2 vectors of time invariant covariates, and $(\beta'_1, \beta'_2, \alpha'_1, \alpha'_2) \in \mathbb{R}^{K_1} \times \mathbb{R}^{K_2} \times \mathbb{R}^{L_1} \times \mathbb{R}^{L_2}$ are unobserved parameters which must be estimated. In this model, $v_{it} \sim i.i.d.(0, \sigma_v^2)$ is independent of $u_i \sim i.i.d.(0, \sigma_u^2)$, so that the variance covariance matrix for the NT observations equals,

$$\begin{aligned} \Sigma &= diag(\sigma_u^2) \otimes J_T + diag(\sigma_v^2) \otimes I_T \\ &= diag(\sigma^2) \otimes \frac{J_T}{T} + diag(\sigma_v^2) \otimes (I_T - \frac{J_T}{T}) && (2.3) \end{aligned}$$

where J_T is a $T \times T$ matrix of ones, I_T is a $T \times T$ identity matrix and $\sigma^2 = T\sigma_u^2 + \sigma_v^2$. In the model, while $E(v_{it}|x_{1,it}, z_{1,i}, x_{2,it}, z_{2,i}) = 0$ and $E(u_i|x_{1,it}, z_{1,i}) = 0$ the remaining covariates are endogenous to u_i; that is $E(u_i|x_{2,it}, z_{2,i}) \neq 0$. Efficient and consistent estimates of the parameters α and β may be obtained via instrumental variable estimation of,

$$\Sigma^{-1/2}y = \Sigma^{-1/2}X_1\beta_1 + \Sigma^{-1/2}X_2\beta_2 + \Sigma^{-1/2}Z_1\alpha_1 + \Sigma^{-1/2}Z_2\alpha_2 + \Sigma^{-1/2}\varepsilon \tag{2.4}$$

where

$$\Sigma^{1/2} = I_{NT} + (1-\theta)P_D \tag{2.5}$$

D stand for the $NT \times N$ matrix of individual dummies, P_D is its projection matrix and $\theta = \left[\sigma_v^2/(\sigma_v^2 + T\sigma_u^2)\right]^{1/2}$. The the instrument set may be built using sample information by exploiting the above distributional assumptions. As usual, the exogenous covariates X_1 and Z_1 provide $K_1 + L_1$ instruments; in addition to this, time invariance provides another $K_1 + K_2$ instruments. To see this last point, $M_D = I_{NT} - P_D = I_{NT} - D(D'D)^{-1}D'$ is its projection matrix, so that M_D transforms any matrix in deviations with respect to the group means. Then, $M(X_1 \vdots X_2)$ provides $K_1 + K_2$ instruments, because $cov(u, M(X_1 \vdots X_2)) = 0$, since $Mu = 0$, although $M(X_1 \vdots X_2)$ is correlated with $(X_1 \vdots X_2)$. Given the $K_1 + K_2 + L_1 + L_2$ covariates and $2K_1 + K_2 + L_1$ instruments, we see that identification is attained provide that $K_1 \geq L_2$. Thus, the final instrument set is given by

$$A = (x_{1,it}; z_{1,i}; x_{1,it} - \bar{x}_{1,i}; x_{2,it} - \bar{x}_{2,i}) \tag{2.6}$$

and its projection matrix is $P_A = A(A'A)^{-1}A'$. Further efficiency gains are available by imposing slightly stronger distributional assumptions. Amemiya and MaCurdy (1986), and Breusch et al. (1989) and Breusch et al. (1999) discuss the conditions under which these assumptions lead to non-redundant additional instruments that improve the accuracy of the estimates (see also Cornwell and Ruppert (1988) and Baltagi and Khanti-Akom (1990)). Although we do not make explicit reference to these alternative contributions in this article, our results carry over in a straightforward fashion.

A drawback of the above estimation method is that the error term is assumed to be homoskedastic, however this is generally not satisfied in the typical application. Therefore, the assumption is now relaxed, by allowing u_i to be independent random variables such that

$$var(u_i|w_i) = \gamma(w_i) = \sigma_{u,i}^2 \tag{2.7}$$

where (hereafter) w_i is a $d \times 1$ vector of (time invariant) informative variables, and might include some or all of the elements in $\bar{x}_{k,i} \otimes \iota_T$, or $z_{k,i}$, for $k = 1, 2$, where $\bar{x}_{k,i}$ has typical element $T^{-1}\sum_t x_{it}$ and ι_T is a $T \times 1$ vector of ones. The mapping $\gamma(.)$ is unknown. The covariance matrix of the regression model is now

$$\Sigma = diag(\sigma_i^2) \otimes \frac{J_T}{T} + diag(\sigma_v^2) \otimes (I_T - \frac{J_T}{T}) \tag{2.8}$$

where $\sigma_i^2 = T\sigma_{u,i}^2 + \sigma_v^2$ varies across i. It is well know (see, for example Hausman and Taylor (1981)) that

$$\Sigma^{-1/2} = I_{NT} - diag(1-\theta_i)P_D. \tag{2.9}$$

When all covariates are exogenous, estimation of β and α could be done via the efficient Random Effects estimator suggested by Roy (2002); however, such a procedure does not seem appropriate here, given the assumed non-zero correlations between Q and u. As alternative procedure can be devised as follows.

Firstly, we can estimate (2.1) via Fixed Effects, obtaining consistent estimates of β; the residual of this regression, say $\hat{v}_{it} - \hat{v}_{i.}$, can then be used to construct a consistent estimator of σ_v^2, namely

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (\hat{v}_{it} - \hat{v}_{i.})^2}{N(T-1) - K}. \tag{2.10}$$

Under the maintained assumptions 2SLS estimation of (2.1) with the instrument set A produces consistent, but inefficient estimates of $\gamma' = (\beta_1', \beta_2', \alpha_1', \alpha_2')$. The residual from this regression, $\tilde{\varepsilon}_{it} = y_{it} - q_{it}'\hat{\gamma}$ compounds the two error components u and v, and under the assumptions imposed on the generating process, it may be used to consistently estimate the conditional moment $E(\varepsilon_{it}^2|w_i) = \sigma_{u,i}^2 + \sigma_v^2$ via kernel regression method[1]. Estimation is discussed below, so let us assume by now that such estimator is available, and denote it by $\tilde{E}(\tilde{\varepsilon}_{it}^2|w_i)$. Then the difference

$$\tilde{\sigma}_{u,i}^2 = \tilde{E}(\tilde{\varepsilon}_{it}^2|w_i) - \hat{\sigma}_v^2. \tag{2.11}$$

provides a consistent estimator for the variance of the individual effects. Substituting $\sigma_{u,i}^2$ and σ_v^2 by their estimates, we obtain a consistent estimator of $\Sigma^{-1/2}$, say $\hat{\Sigma}^{-1/2}$. Finally, consistent and efficient estimates of the parameters in the model will follow by implementing the generalize 2SLS estimation of (2.4) with the set of instruments A and $\hat{\Sigma}^{-1/2}$ replacing $\Sigma^{-1/2}$. That is,

$$\hat{\gamma}^{(1)} = (Q'\hat{\Sigma}^{-1/2}P_A\hat{\Sigma}^{-1/2}Q)^{-1}Q'\hat{\Sigma}^{-1/2}P_A\hat{\Sigma}^{-1/2}Y \tag{2.12}$$

$$\hat{\Sigma}^{-1/2} = I_{NT} + diag(1-\hat{\theta}_i)P_D \tag{2.13}$$

and $\hat{\theta}_i^2 = \hat{\sigma}_v^2/(\hat{\sigma}_v^2 + T\hat{\sigma}_{u,i}^2)$.

In practice researchers will often be interested in evaluating instrument relevance after estimation. Then, the measures in Shea (1997) and Godfrey (1999) can be applied directly but bearing in mind that comparisons show now involve the ratio of variances and error variances of the IV method just described with respect to the efficient GLS method in Roy (2002).

2.1 Nonparametric Estimation

The above procedures require the use of a consistent estimators for the moment $E(\varepsilon_{it}^2|w_{it})$. These are readily available from the literature on nonparametric methods (see, for instance, surveys by Silverman (1986), Hardle (1990), or more recent and comprehensive contributions by Pagan and Ullah (1999) and Li and Racine (2006)). The Random Effects models in Roy (2002) and Li and Stengos (1994) use a Nadaraya-Watson regression (NW) to estimate the aforementioned moment. This estimator solves the problem,

$$\min_m(x) \sum_k (y_k - m(x))^2 K_h(\frac{x_k - x}{h}). \tag{2.14}$$

This is as a weighted least squares problem where $K(.)$ is the usual kernel function[2], $h = h(n)$ is the bandwidth parameter (which satisfies $h \to 0$, $nh^d \to \infty$ as $n \to \infty$), $m(x) = E(y|x)$ and,

[1] Alternatively, series and spline regression methods are available.

[2] For simplicity, K_h is assumed to be a product kernel, so that, if x has dimension q, then $K_h(x) = \frac{1}{h^q}\prod_{j=1}^q k(x_j)$, where $k(.) : \mathbb{R} \mapsto \mathbb{R}^+$

therefore, $\hat{m}(x)$ is a local estimator of $m(x)$ at x. The NW estimator is simple to compute, consistent and asymptotically normal, thus facilitating inference. However, it is well known that use of local polynomial regression (LPR) to estimate conditional moments can yield estimates of a better quality. There are two key reason underlying this claim. Firstly LPR solves the alternative least squares problem

$$\min_{\mathbf{m}(x)} \sum_{k}(y_k - m(x) - m^{(1)}(x)(x_k - x) \dots m^{(p)}(x)(x_k - x)^p)^2 K_h(\frac{x_k - x}{h}) \tag{2.15}$$

where $\mathbf{m}(x) = (m(x), m^{(1)}(x), \dots, m^{(p)}(x))'$. In above optimization problem, the variable of interest is regressed on a p^{th} order Taylor Series approximation of the unknown form of the conditional moment ($m^{(p)}$ denotes the p^{th} partial derivative of $m(.)$ with respect to x). From this point of view, the NW relies on a zero order approximation of the unknwn moment, and therefore, LPR are likely to reduce the amount of variability due to approximation errors. Secondly, LPR has the potential to mitigate the boundary bias problem. At the boundary of X , the NW's error of estimation disappears asymptotically, but at a rate slower than in the interior of the set on which the regressors take values from, so that larger amounts of data are required in the boundaries of the sample to obtain an error of magnitude comparable to that in the interior of the set of regressors. In the typical application, however, sample points are sparse in the boundaries, and this claims for boundary bias corrections when implementing kernel methods in small samples. It is known (see, for example, Wand and Jones (1994)) that the rate of convergence of LPR is similar to that of NW, however this rate is uniform across the range of X, thus improving the quality of the estimates in those areas in finite samples. Given these two features of the estimators, this article explores if estimating the conditional second moments via LPR as opposed to NW brings further efficiency gains in estimation.

3 Monte Carlo

In order to justify our claims of efficiency gains we report the results of a Monte Carlo experiment. The Data Generating Processes used in the experiment combined specifications in Im et al. (1999) and Roy (2002) to allow for endogenous covariates and heteroskedastic individual effects. To produce a panel with heteroskedastic random effects, data was drawn from the linear model

$$y_{it} = \beta_0 + x'_{it}\beta + z'_i\gamma + u_i + v_{it}.$$

where $x_{it} = (x_{1,it}, x_{2,it}, x_{3,it}, x_{4,it})'$, $z_i = (z_{1,i}, z_{2,i})'$ $\beta_j = \gamma_j = 1$. For $j = 1, 2$, $x_{j,it} = 0.7x_{j,it-1} + \delta_{j,i} + \epsilon_{j,it}$. For $j = 3, 4$ $x_{j,it} = 0.7x_{j,it-1} + \delta_{j,i} + u_i + \epsilon_{j,it}$, so that these covariates are endogenous. The coefficient 0.7 ensures these processes are stationary. For $t = 0$, $x_{j,it} = 0$. All $\delta_{j,i}$, $\epsilon_{j,it}$ are drawn from Uniform distributions on $(-2, 2)$. Similarly, $z_{1,i} = 0.5 * \delta_{1,i} + 0.5\delta_{2,i} + \xi_{1,i}$, while $z_{2,i} = \delta_{1,i} + \delta_{2,1} + u_i + \xi_{2,i}$. Here $\xi_{j,i}$ are uniformly distributed on $(-2, 2)$. In this specification, $z_{1,i}$ is correlated with $x_{1,it}, x_{2,it}$ via $\delta_{1,i}$ and $\delta_{2,i}$, but uncorrelated with u_i and v_{it}. On the other hand, $z_{2,i}$ is correlated with $x_{1,it}, x_{2,it}$ as well as u_i.

The error components were $v_{it} \sim N(0, \sigma_v^2)$, independent of $u_i \sim N(0, \omega_i)$, where $\omega_i = \alpha^2(1 + \lambda z_{1,i})^2$. The total variance was fixed to $E(v_{it}^2) + E(\omega_i) = 8$. The contribution of σ_v^2 was controlled in the simulations, by assigning values 2, 4 and 6, corresponding to 25%, 50% and 75% of the total expected variance, while the functional form of the heteroskedasticity was ruled by the parameter λ. This parameter was assigned values 0 (corresponding to homoskedastic individual effects), 1, 2 and 4 (with larger values increasing the curvature of ω_i. Each choice of values for the pair $(\sigma_v^2; \lambda)$ implies a value of α, which then can be derived by solving the equation for the expected total variance.

We used Ox V.5. for the simulations[3]. The results reported here are based on $R = 2000$ replication. We report the average standard error of each parameter as well as the magnitude $MSR = R^{-1}\sum_r(\hat{\beta}_{j,r} - \beta_{j,0})^2$, where $\beta_{j,0}$ is the true value of the parameter. Without loss of generality[4], here we report results for $N = 50, T = 5$, $\sigma_v^2 = 4$. These results are collected in Tables 1 to 3.

As in Roy (2002) and Baltagi et al. (2005) , we find that the results are fairly robust to the choice of bandwidth, at leat for this simple case where only one covariate has been considered in the nonparametric regressions. In gereral, all methods exhibit a loss of efficiency in the estimation of the endogenous variables with respect to the exogenous variables, and the coefficeints of time varying covariates are estimated much more efficienctly than the coefficients of time invariant ones. Simulations not reported here show that as σ_v^2 increases, so does the mean standard error and MSE of the estimators, while increases in the sample size mitigate this loss of efficiency.

For the case $\lambda = 0$ (homoskedasticity), all four estimators provide very similar results. The HT estimator seems to dominate all the others, as was expected. However, on occassions the adaptive method can be more efficient than HT. This is most noticeable with the time invariant parameters and the intercept, where and adaptive estimator of based on a LPR of order 2 is more efficient than HT. Once heteroskedasticity is introduced in the generating process ($\lambda > 0$), the adaptive methods exhibit a clear comparative advantage. Although efficiency gains can be observed for all the parameters, the use of adaptive estimators has its most notorious effect on the coefficients of the time invariant variables. Efficiency gains for these parameters are in the region of a 12% reduction on average s.d. and a 20% reduction on the MSE with respect to the estimator in Hausman and Taylor.

Within the group of adaptive estimators, we observe some efficiency gains when a local linear regression is used in place of the Nadaraya-Watson estimator, and often local polynomial regression of order 2 and 3 could provide some efficiency gains. However, these efficiency gains are not consistent in our results, and they seem to vary from case to case, which ultimately suggests there are not net comparative advantages in the use of LPR of orders above 1. On ocassions, these might even induce a small efficiency loss with respect to the adaptive estimator that uses the Nadaraya Watson regression.

Tables 4 and 5 collect the empirical size and power of the t-ratios. We only report the case $\sigma_v^2 = 4$ for the Hausman Taylor estimator and adaptive estimators using a Nadaraya Watson and Local Linear Regression estimators of the conditional variances of the error components. Attention is focused on the coefficients β_4 and α_2 corresponding to a time varying and time invariant endogenous regressor, as well as β_2 (a exogenous time varying regressor) for comparison. There is no major size distortion as far as the t-ratio for β_2 is concerned; however the size distortion of the t-ratio when testing $\beta_4 = 0$ is considerable (exceeding 5%) and $\alpha_2 = 0$ (well below the 5% significance level). The performance of the test is more or less constant for the adaptive estimators across different patterns of heteroskedasticity, however heteroskedasticity seems to affect the t-ratios for α_2 derived from the Hausman-Taylor estimator. In this case, the empirical size falls even further for large λ. The t-ratio succeeds to reject the null hypothesis when falls 100% of the time as far as the coefficients of the time-varying covariates is concerned. These tests exhibit low power when testing the coefficients of time-invariant covariates. This loss of power is mitigated if inference is based on the adaptive estimator, provided that $\lambda > 0$.

Overall the adaptive Hausman Taylor estimators provide significant efficiency gains in estimation of a linear error components model when the individual effects are heteroskedastic. This efficiency gain affects primarily the coefficients of the time-invariant covariates and the corre-

[3]The codes are available for inspection upon request.

[4]For different choices of these parameters, the absolute value of the magnitudes changed, but the conclusions stayed the same. The un-tabulated results of the whole simulation are, of course, available upon request.

sponding t-ratios. Among the class of adaptive estimators, we observe modest efficiency gains if local linear regression is used to estimate the conditional variance of the compound error, as opposed to the simpler Nadaraya-Watson. However, it is unclear whether these gains justify the added complexity in the model.

4 Conclusion

This note presents a class of adaptive IV of estimator of a error component model with heteroskedastic random effects. The estimator is inspired on earlier work by Roy (2002), and is a natural extension of Hausman and Taylor (1981), although the approach applies without modifications to the estimators found in Amemiya and MaCurdy (1986) and Breusch et al. (1989)). The method relies on nonparametric regression to estimate the conditional variance of the compound error term, and we explored whether different estimators lead to substantial differences in the results. The new method provides significant efficiency gains with respect to the original estimator and it is useful in a variety of cases, such as the estimation of Stochastic Frontiers where firm's inefficiency, which is is captured through the time invariant random effect, is often expected to be differently distributed across clusters of firms.

References

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6:21–37.

Amemiya, T. and MaCurdy, T. (1986). Instrumental-variable estimation of an error-components model. *Econometrica*, 54:869–880.

Baltagi, B., Bresson, G., and Pirotte, A. (2005). Adaptive estimator of heteroskedastic error component models. *Econometric Reviews*, 24:39–58.

Baltagi, B. and Griffin, J. (1988). A generalized error component model with heteroscedastic disturbances. *International Economic Review*, 29:745–753.

Baltagi, B. and Khanti-Akom, S. (1990). On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, 5:401–406.

Breusch, T., Mizon, G., and Schmidt, P. (1989). Efficient estimation using panel data. *Econometrica*, 57:695–700.

Breusch, T., Qian, H., Schmidt, P., and Wyhowski, D. (1999). Redundancy of moment conditions. *Journa of Econometrics*, 91:89–111.

Cornwell, C. and Ruppert, P. (1988). Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, 3:149–155.

Godfrey, L. (1999). Instrument relevance in multivariate linear models. *The Review of Economics and Statistics*, 81:550–552.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

Hausman, J. and Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica*, 49:1377–1398.

Im, K., Ahn, S., Schmidt, P., and Wooldridge, J. (1999). Efficient estimation of panel data models with strictly exogenous explanatory variables. *Journa of Econometrics*, 93:177–201.

Li, Q. and Racine, J. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Li, Q. and Stengos, T. (1994). Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. *International Economic Review*, 35:981–1000.

Mazodier, P. and Trognon, A. (1978). Heteroskedasticity and stratification in error components models. *Annales de l'INSEE*, 30-31:451–482.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Roy, N. (2002). Is adaptive estimation useful for panels with heteroskedasticity in the individual specific error component? some monte carlo evidence. *Econometric Reviews*, 21:189–203.

Schmidt, P. and Sickles, R. (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics*, 2:367–374.

Shea, J. (1997). Instrument relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics*, 79:348–352.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

Wand, M. and Jones, M. (1994). *Kernel smoothing.* Chapman and Hall/CRC.

N=50, T=5
Bandwidth $h = 0.2$.

DGP $\lambda = 0$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10472	0.01061	0.10385	0.01073	0.10406	0.01071	0.10353	0.01142	0.10374	0.01186
β_2	0.10443	0.01083	0.10359	0.01091	0.10336	0.01004	0.10304	0.01048	0.10371	0.01130
β_3	0.12453	0.01698	0.12429	0.01693	0.12520	0.01559	0.12489	0.01558	0.12490	0.01632
β_4	0.12477	0.01698	0.12453	0.01689	0.12502	0.01527	0.12485	0.01620	0.12511	0.01627
β_0	0.43251	0.18423	0.41270	0.19545	0.41490	0.15970	0.40431	0.15778	0.40613	0.15261
α_1	0.40768	0.16479	0.38739	0.16534	0.39523	0.15783	0.37466	0.13271	0.37546	0.16354
α_2	0.53565	0.29709	0.52114	0.30439	0.54006	0.31137	0.50635	0.26338	0.50878	0.30186

DGP $\lambda = 1$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10377	0.01049	0.10058	0.00998	0.10142	0.01029	0.09988	0.00972	0.10040	0.00990
β_2	0.10362	0.01026	0.10035	0.00980	0.10087	0.00976	0.10135	0.00975	0.10125	0.01008
β_3	0.12547	0.01544	0.12566	0.01554	0.12584	0.01605	0.12543	0.01615	0.12559	0.01677
β_4	0.12547	0.01671	0.12564	0.01685	0.12595	0.01584	0.12527	0.01595	0.12568	0.01638
β_0	0.41952	0.15004	0.36498	0.11678	0.37245	0.12067	0.36523	0.12296	0.36268	0.11668
α_1	0.39082	0.15671	0.34445	0.10978	0.36057	0.11056	0.34812	0.11410	0.35166	0.11537
α_2	0.50753	0.23536	0.44327	0.18321	0.45825	0.20682	0.44073	0.18678	0.44020	0.18039

DGP $\lambda = 2$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10418	0.010624	0.10080	0.01010	0.10113	0.01056	0.10105	0.01061	0.10083	0.01007
β_2	0.10420	0.010108	0.10098	0.00974	0.10105	0.01057	0.10058	0.01017	0.10069	0.00990
β_3	0.12557	0.015617	0.12576	0.01565	0.12537	0.01583	0.12553	0.01598	0.12603	0.01666
β_4	0.12538	0.015924	0.12555	0.01601	0.12554	0.01596	0.12555	0.01531	0.12604	0.01565
β_0	0.42403	0.15296	0.35246	0.10341	0.35748	0.10620	0.34945	0.09667	0.34993	0.09828
α_1	0.40065	0.16916	0.36353	0.13313	0.36731	0.12766	0.36317	0.12939	0.35885	0.12443
α_2	0.52162	0.28568	0.44400	0.19353	0.44589	0.19884	0.44051	0.18861	0.43749	0.18137

Table 1: Results from Simulation. $R = 2000$

N=50, T=5
Bandwidth $h = 0.4$.

DGP $\lambda = 0$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10472	0.01061	0.10334	0.01081	0.10287	0.01068	0.10330	0.01028	0.10390	0.01091
β_2	0.10443	0.01083	0.10389	0.01049	0.10422	0.01060	0.10407	0.01016	0.10406	0.01212
β_3	0.12453	0.01698	0.12484	0.01584	0.12497	0.01672	0.12486	0.01586	0.12511	0.01568
β_4	0.12477	0.01698	0.12483	0.01632	0.12514	0.01696	0.12508	0.01604	0.12507	0.01695
β_0	0.43251	0.18423	0.41771	0.15846	0.40197	0.14719	0.41136	0.18772	0.42198	0.19338
α_1	0.40768	0.16479	0.39180	0.18137	0.37730	0.14399	0.36792	0.12002	0.39365	0.16691
α_2	0.53565	0.29709	0.52929	0.31701	0.50973	0.26593	0.50446	0.24380	0.53364	0.34394

DGP $\lambda = 1$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10377	0.01049	0.10105	0.01003	0.10091	0.01003	0.10068	0.00995	0.10105	0.01004
β_2	0.10362	0.01026	0.10049	0.01017	0.10120	0.01008	0.10080	0.00959	0.10106	0.00991
β_3	0.12547	0.01544	0.12593	0.01657	0.12588	0.01714	0.12617	0.01710	0.12598	0.01598
β_4	0.12547	0.01671	0.12601	0.01578	0.12567	0.01567	0.12580	0.01661	0.12573	0.01577
β_0	0.41952	0.15004	0.37335	0.11853	0.36946	0.11263	0.37012	0.12396	0.36667	0.12198
α_1	0.39082	0.15671	0.35526	0.11698	0.35260	0.11337	0.35597	0.12429	0.35192	0.11173
α_2	0.50753	0.23536	0.44733	0.18618	0.44917	0.20457	0.44919	0.20145	0.44334	0.19295

DGP $\lambda = 2$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10418	0.01062	0.10000	0.00976	0.10066	0.01029	0.10077	0.01014	0.10123	0.01063
β_2	0.10420	0.01010	0.10094	0.00994	0.10085	0.01027	0.10133	0.01030	0.10065	0.01000
β_3	0.12557	0.01561	0.12590	0.01602	0.12585	0.01610	0.12603	0.01737	0.12587	0.01713
β_4	0.12538	0.01592	0.12558	0.01618	0.12602	0.01613	0.12602	0.01672	0.12591	0.01612
β_0	0.42403	0.15296	0.34759	0.09826	0.35368	0.10568	0.35191	0.10249	0.35616	0.10168
α_1	0.40065	0.16916	0.36232	0.13227	0.36785	0.13609	0.36541	0.14172	0.37044	0.13616
α_2	0.52162	0.28568	0.43481	0.17758	0.43786	0.18980	0.43937	0.18498	0.45284	0.19192

Table 2: Results from simulation. $R = 2000$

N=50, T=5
Bandwidth $h = 0.8$.

DGP $\lambda = 0$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10472	0.01061	0.10447	0.01110	0.10339	0.01151	0.10353	0.01142	0.10334	0.01029
β_2	0.10443	0.01083	0.10351	0.01026	0.10346	0.01100	0.10304	0.01048	0.10404	0.01016
β_3	0.12453	0.01698	0.12540	0.01642	0.12517	0.01563	0.12489	0.01558	0.12531	0.01632
β_4	0.12477	0.01698	0.12496	0.01639	0.12532	0.01556	0.12485	0.01620	0.12498	0.01605
β_0	0.43251	0.18423	0.41924	0.16640	0.40212	0.14301	0.40431	0.15778	0.43950	0.17118
α_1	0.40768	0.16479	0.38449	0.16483	0.37106	0.14888	0.37466	0.13271	0.42330	0.17982
α_2	0.53565	0.29709	0.52708	0.29527	0.50126	0.27860	0.50635	0.26338	0.56610	0.31250

DGP $\lambda = 1$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10377	0.01049	0.10064	0.00958	0.10129	0.00969	0.099881	0.00972	0.10085	0.01053
β_2	0.10362	0.01026	0.10061	0.00979	0.10107	0.00990	0.10135	0.00975	0.10130	0.01043
β_3	0.12547	0.01544	0.12541	0.01624	0.12608	0.01675	0.12543	0.01615	0.12593	0.01664
β_4	0.12547	0.01671	0.12596	0.01648	0.12616	0.01617	0.12527	0.01595	0.12598	0.01553
β_0	0.41952	0.15004	0.36686	0.11409	0.37868	0.12616	0.36523	0.12296	0.36559	0.11807
α_1	0.39082	0.15671	0.34618	0.11239	0.36252	0.12053	0.34812	0.11410	0.35175	0.11757
α_2	0.50753	0.23536	0.43981	0.17522	0.45376	0.20188	0.44073	0.18678	0.44375	0.19901

DGP $\lambda = 2$, $\sigma_v = 4$

	HT		Polynomial 0		Polynomial 1		Polynomial 2		Polynomial 3	
	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE	S.d.	MSE
β_1	0.10418	0.01062	0.10123	0.00994	0.10084	0.01018	0.10105	0.01061	0.10050	0.01048
β_2	0.10420	0.01010	0.10099	0.01002	0.10081	0.01001	0.10058	0.01017	0.10130	0.01086
β_3	0.12557	0.01561	0.12545	0.01614	0.12587	0.01662	0.12553	0.01598	0.12556	0.01647
β_4	0.12538	0.01592	0.12579	0.01631	0.12604	0.01714	0.12555	0.01531	0.12537	0.01650
β_0	0.42403	0.15296	0.35259	0.10636	0.35300	0.11061	0.34945	0.09667	0.35078	0.10578
α_1	0.40065	0.16916	0.36483	0.13791	0.36572	0.13842	0.36317	0.12939	0.35442	0.12629
α_2	0.52162	0.28568	0.44505	0.22778	0.43516	0.19059	0.44051	0.18861	0.45244	0.19733

Table 3: Results from simulation. $R = 2000$

Empirical Significance Levels t-ratios (Size)

	$N=50,\ T=5$			$N=100,\ T=5$		
	DGP $\lambda=0,\ \sigma_v=4$; Bandwidth $h=0.2$.					
	HT	NW	LLR	HT	NW	LLR
β_2	4.10	4.40	4.20	5.15	4.20	4.10
β_4	5.60	5.70	5.60	4.90	5.30	5.40
α_2	2.40	3.00	3.15	4.20	3.65	3.60
	DGP $\lambda=1,\ \sigma_v=4$; Bandwidth $h=0.2$.					
	HT	NW	LLR	HT	NW	LLR
β_2	4.50	4.85	4.85	5.00	5.15	5.00
β_4	6.55	6.75	6.65	4.05	6.05	6.05
α_2	2.70	3.05	3.00	3.60	4.70	4.50
	DGP $\lambda=2,\ \sigma_v=4$; Bandwidth $h=0.2$.					
	HT	NW	LLR	HT	NW	LLR
β_2	5.80	6.40	6.25	4.05	4.85	4.75
β_4	5.55	5.70	5.75	5.60	5.75	5.70
α_2	3.60	4.30	4.30	2.85	4.20	4.05

Table 4: Proportion of Rejections of the Null Hypothesis ($H_o : \beta_j = 0$)

Proportion of Rejections t-ratios

	$N=50,\ T=5$			$N=100,\ T=5$		
	DGP $\lambda=0,\ \sigma_v=4$; Bandwidth $h=0.2$.					
	HT	NW	LLR	HT	NW	LLR
β_2	100.00	100.00	100.00	100.00	100.00	100.00
β_4	100.00	100.00	100.00	100.00	100.00	100.00
α_2	70.65	63.90	65.20	81.25	82.00	82.55
	DGP $\lambda=1,\ \sigma_v=4$; Bandwidth $h=0.2$.					
	HT	NW	LLR	HT	NW	LLR
β_2	100.00	100.00	100.00	100.00	100.00	100.00
β_4	100.00	100.00	100.00	100.00	100.00	100.00
α_2	64.40	69.85	66.00	83.05	89.80	89.90
	DGP $\lambda=2,\ \sigma_v=4$; Bandwidth $h=0.2$.					
	HT	NW	LLR	HT	NW	LLR
β_2	100.00	100.00	100.00	100.00	100.00	100.00
β_4	100.00	100.00	100.00	100.00	100.00	100.00
α_2	62.45	67.65	66.45	83.65	90.90	91.10

Table 5: Proportion of Rejections of the Null Hypothesis ($H_o : \beta_j = 0$)