Heteroskedasticity and Autocorrelation Robust Inference for a System of Regression Equations

Robert D. J. Anderson [a], Ralf Becker [b], Denise R. Osborn [b]

a Business School - Economics, Newcastle University, NE1 7RU, England, UK
b Economics, University of Manchester, M15 5PU, England, UK

No 2009/11

Newcastle Discussion Papers in Economics: ISSN 1361 - 1837

Heteroskedasticity and Autocorrelation Robust Inference for a System of Regression Equations

November 2009

Robert D. J. Anderson [a], **Ralf Becker** [b], **Denise R. Osborn** [b]
[a] Business School - Economics, Newcastle University, NE1 7RU, England, UK
[b] Economics, University of Manchester, M15 5PU, England, UK

Abstract

This paper extends standard single equation heteroskedasticity and autocorrelation (HAC) robust inference methods to allow consistent inference for a system of vector moving-average correlated equations also accommodating contemporaneous correlations. This is of particular relevance to the examination of inflation forecast errors, as forecasts for different groups are contemporaneously correlated, while any proposed forecasting model utilising a time-series of multi-period forward-looking expectations data will suffer from overlapping errors inducing a moving-average error structure. The proposed methodology is a generalisation of Newey & West (1987) and the SUR technique of Zellner (1962). Monte Carlo simulations confirm that the method performs well in large samples. Applications testing the rationality of male versus female inflation forecasts, and those of defined educated groups, are also included.

We are grateful to the helpful suggestions provided by members of the Econometrics and Applied Economics research area group at the University of Manchester and also members of the Economics division of the Business School at Newcastle University. The usual disclaimer applies that all errors and omissions are entirely the responsibility of the authors.

1 Introduction

Underpinning the validity of all applied econometric analysis are the assumptions made by the researcher regarding the underlying data generation process. For example, is the process stationary, does it contain a unit root, are the errors independently and identically distributed (spherical), or is there some degree of dependency between errors. Importantly, where a researcher fails to sufficiently account for any dependencies in the data, then estimated regression models may provide misleading inference, with corresponding coefficient standard-errors producing test-statistics which are incorrectly sized, and/or have test power which is severely reduced.

Methods to improve inference efficiency given the failure of classical OLS assumptions have generally been developed either in the context of the analysis of time-series data or for the analysis of survey-data. Relatively few methods exist for dealing with a mixture of both data types, henceforth referred to as time-series survey-data (TSSD)[1] which is distinctive in having a large number of time-periods compared to the number of panels or cross-sections. TSSD can be investigated by treating each repeated panel or cross-section as one time-series equation in a system-of-equations (SoE). Estimation of such a system taking account of the correlation structure of the system can be problematic, as two-dimensional dependency may exist *within* and *between* equation disturbances. Further, using standard (time-series or cross-section) robust inference techniques in this context may produce misleading inference since, implicitly, only correlations in one-dimension are being accommodated.

To our knowledge, no non-parametric[2] method has been developed to date which will produce consistent inference for parameters in a SoE when the form of the two-dimensional correlation process is unspecified. In particular, no method is specifically designed to accommodate a (vector) moving average error process combined with possible heteroskedasticity within the system equation error processes. This paper proposes such a method which is both applicable to the situation of a vector moving average error process with contemporaneous cross-equation correlations with or without heteroskedasticity, and is robust to more general forms of system error correlation structure.

---

[1] Other definitions used in the literature for this type of data include a ‘long-wide’ dataset, and ‘time-series cross-section data’. The latter may misleadingly suggest that the sampled units are necessarily different over time, which need not be the case.

[2] Not involving direct-estimation of the parameters of the error process.

The remainder of the paper is structured as follows: section 2 reviews the rationale for using a SoE and discusses methods currently available to deal with various error-dependency structures. Having introduced appropriate notation, section 3 presents the new methodology with reference to accommodating a system vector-moving average error structure. Simulations are conducted to assess the finite sample size and power properties of this new methodology, the results being presented in section 4. Section 5 presents an application of the new methodology, with final conclusions being drawn in section 6.

2 Time-Series Survey Data Regressions

Consistent inference for a TSSD equation system using current econometric methodologies implicitly requires the researcher to make, sometimes strong, assumptions regarding the correlation structure of the system errors. Using a methodology which fails to fully accommodate possible correlation structures will result in inefficient or incorrect inference. This section reviews current methodologies, having first introduced the SoE notation.

Consider a set of linear regression models, containing equations (or groups) $i=\{1,2,\ldots n\}$ each repeatedly sampled over time $t=\{1,2,\ldots T\}$[3]. Each equation has k_i exogenous explanatory variables ($\sum_{i=1}^{n} k_i = K$ in total), written compactly as:

$$\underset{T\times 1}{Y_i} = \underset{T\times k_i}{X_i}\,\underset{k_i\times 1}{\boldsymbol{\beta}_i} + \underset{T\times 1}{u_{i\cdot}} \tag{1}$$

OLS estimation of each of these individual equations (as opposed to the entire system) is referred to as equation-by-equation OLS. The full system can be defined as:

$$Y = X\beta + u \tag{2}$$

where:

$$\underset{nT\times 1}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{nT\times K}{X} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{bmatrix} \quad \underset{K\times 1}{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix} \quad \underset{nT\times 1}{u} = \begin{bmatrix} u_{1\cdot} \\ u_{2\cdot} \\ \vdots \\ u_{n\cdot} \end{bmatrix} \tag{3}$$

The structure of the regressor matrix is unconstrained. However the block-diagonal structure in (3) will produce coefficient estimates which only measure effects *within* equations. If desired, a simple reparameterisation of the regressor matrix, as demonstrated by equation (4), allows the estimated coefficients to measure the difference *between* a specific (base) equation, in this case the first, and the other equations.

$$\underset{nT\times K}{X} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ X_2 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ X_n & 0 & \cdots & X_n \end{bmatrix} \quad \underset{K\times 1}{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n - \boldsymbol{\beta}_1 \end{bmatrix} \tag{4}$$

---

[3] The lack of equation time subscript implies that the system is balanced: all equations are sampled over the same time period, the same number of times, at the same time-interval (monthly, quarterly, etc.). The requirement that the sampling interval is constant for over all *t* is important when considering a bandwidth size, to be discussed shortly.

In its most general form, the system-error covariance matrix has the form:

$$E\left(uu'\right)=\underset{nT\times nT}{\Omega}=\begin{bmatrix}\sigma_{11}M_{11} & \sigma_{12}M_{12} & \cdots & \sigma_{1n}M_{1n}\\ \sigma_{21}M_{21} & \sigma_{22}M_{22} & \cdots & \sigma_{2n}M_{2n}\\ \vdots & \vdots & & \vdots\\ \sigma_{n1}M_{n1} & \sigma_{n2}M_{n2} & \cdots & \sigma_{nn}M_{nn}\end{bmatrix} \tag{5}$$

Within equation heteroskedasticity and autocorrelation is captured by the equation (and cross-equation) specific square $\left(T\times T\right)$ M_{ij} matrices[4]. Heteroskedasticity and contemporaneous-correlations across equations are captured through the equation and cross-equation (co-)variance scalars σ_{ij} calculated using equation-by-equation OLS residuals as:

$$\hat{\sigma}_{ij}=\frac{\hat{u}_{i\cdot}{}'\hat{u}_{j\cdot}}{\sqrt{T-k_i}\sqrt{T-k_j}}\quad \text{for } i,j=1,\ldots,n \tag{6}$$

A commonly used technique to estimate a SoE is the method of Seemingly Unrelated Regressions, SUR (see Zellner 1962), which accommodates contemporaneous-correlations *between* equations. This method does, however, impose the restriction that $M_{ij}=I_T$ (where I_T is an identity matrix of dimension *T*), so ruling out within-equation correlations and within-equation heteroskedasticity. The extension to this methodology proposed by Parks (1967) permits *within*-equation residual autocorrelation, but assumes an autoregressive (AR) residual correlation structure in the construction of the M_{ij} matrices[5], and maintains the assumption of *within*-equation homoskedasticity.

In summary, standard SUR and the extension to SUR allowing for an autoregressive error structure (the Parks method), implicitly use GLS to estimate the coefficients and the corresponding coefficient covariance matrix:

$$\begin{aligned}\hat{\beta}_j&=\left(X'\hat{\Omega}_j^{-1}X\right)^{-1}X'\hat{\Omega}_j^{-1}X\quad \text{for } j=\text{SUR, Parks}\\ Var&=\left(X'\hat{\Omega}_j^{-1}X\right)^{-1}\end{aligned} \tag{7}$$

---

[4] The elements on the main diagonal of the M_{ij} matrix scale the average variance term, σ_{ij}, to the time-specific level required, effectively resulting in $\sigma_{ij,t}$.

[5] Appropriate commands in Stata8.0 designed to accommodate contemporaneous correlations, such as ‘xtgls’ and ‘xtpcse’, restrict the permissible autoregressive autocorrelation structure to being first-order process.

where in each case, the system covariance matrix is constructed as:

$$\hat{\Omega}_{SUR} = \hat{\Sigma} \otimes I_T \tag{8}$$

$$\hat{\Omega}_{Parks} = \hat{P}\hat{\Omega}_{SUR}\hat{P}' \tag{9}$$

where:

$$\hat{\Sigma} = \left\{\hat{\sigma}_{ij}\right\} \tag{10}$$

and $\hat{P}$ is block-diagonal with typical block $\left\{\hat{P}_i\right\}$, with

$$P_i P_j' = M_{ij} = \frac{1}{1-\hat{\rho}_i\hat{\rho}_j}\begin{bmatrix} 1 & \hat{\rho}_j & \cdots & \rho_j^{T-1} \\ \hat{\rho}_i & 1 & \cdots & \rho_j^{T-2} \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_i^{T-1} & \hat{\rho}_i^{T-2} & \cdots & 1 \end{bmatrix} \tag{11}$$

such that (10) represents the contemporaneous system covariance matrix, with elements estimated in the usual way, and $\hat{P}$ represents a transformation matrix accounting for within-equation correlations where $\hat{\rho}_i$ is the parameter estimate from an AR(1) processes in the estimated residuals for equation *i*.

It is known that an (vector) autoregressive (AR) process of infinite order can be used to approximate an invertible (vector) moving average (MA) process of finite order. Galbraith and Zinde-Walsh (1994), and Galbraith *et al* (2002) for the vector case, consider this approximation, and shows that in certain situations, particularly in small samples, the (V)AR representation is biased. Further, they demonstrate that this bias increases the nearer the (V)MA process is to non-invertibility. Approximating a (V)MA process by a low-order (V)AR process would also result in a poor performance, and so a key problem with such an approximation is choosing a suitable order for the (V)AR process, particularly when the order of the (V)MA process is unknown. Clearly, therefore, a (V)AR approximation is not always appropriate or straightforward.

Estimation using the Parks methodology involves two initial steps: (i) estimating the autoregressive parameters from OLS equation-by-equation residuals (ie. constructing the $\hat{P}_i$ matrix), (ii) transforming the residuals for each equation to remove within equation correlations and then estimating the contemporaneous covariance terms (equivalent to constructing $\hat{\Omega}_{SUR}$). Coefficients and associated standard errors are then estimated using feasible-GLS as per (7).

Beck and Katz (1995) show, using Monte Carlo simulations, that in certain situations (notably where $T/n > 3$), inefficiency can be introduced at each step, causing the Parks methodology to generate oversized test statistics for the regression coefficients, and so produce misleading inference. Beck and Katz suggest partially compensating for this bias by using the estimated system-covariance matrix after OLS estimation, rather than using feasible-GLS. They call this method Panel Corrected Standard Errors (PCSE). Therefore:

$$\begin{aligned} \hat{\beta}_{PCSE} &= \left(X'X\right)^{-1} X'Y \\ Var\left(\hat{\beta}_{PCSE}\right) &= \left(X'X\right)^{-1} X'\hat{\Omega}_{PCSE} X \left(X'X\right)^{-1} \end{aligned} \tag{12}$$

When $\hat{\Omega}_{PCSE}$ is constructed as per $\hat{\Omega}_{SUR}$, this is denoted by a PCSE/SUR subscript, which is only permitting *between* equation contemporaneous correlations. On the other hand, when $\hat{\Omega}_{PCSE}$ is constructed as per $\hat{\Omega}_{Parks}$, that is where there are also *within* equation correlations (implicitly of an autoregressive type), this is denoted by a PCSE/Parks subscript.

All these methods utilise an estimated system-covariance matrix, $\hat{\Omega}$, to produce corrected coefficient standard errors, and in the case of GLS based methods, to produce parameter estimates. Constructing such a matrix not only allows immediate (potentially robust) tests of parameter significance but also allows post-estimation tests of both cross- and within-equation restrictions using the robust Wald test (see Wooldridge 2002a or Godfrey and Orme 2003). However, if the within-equation autocorrelation structure is not well approximated by an autoregressive process, or there is within-equation heteroskedasticity, it is clear that current methods will not provide robust consistent estimates. Furthermore, even when a Parks-type method might seem appropriate, the evidence of Beck and Katz suggests that inference may still be unreliable.

Although full parametric transformations involving maximum likelihood, and bootstrapping[6] estimated standard-errors could both be used to address some of the methodological shortcomings (at-least with a single-equation), both these techniques can be problematic to

---

[6] Using a block-bootstrap methodology to preserve the correlation-structure. For further information see MacKinnon (2002)

apply[7]. In a single equation context, it is far more common to use techniques which will consistently estimate an asymptotically (heteroskedasticity and autocorrelation) robust covariance matrix, removing the need to specify a precise model for the form of the data-dependency.

The next section introduces a new methodology, extending the single-equation heteroskedasticity and autocorrelation robust covariance matrix estimation technique of Newey and West (1987) to a SoE, allowing robust estimation of $X'\Omega X$. This technique is particularly useful in this setting as not only does it accommodate possible heteroskedasticity and a wide range of correlation structures, but using a modified Andrews (1991) bandwidth methodology, all parameters of this technique can be made data-dependent. It represents an extension to the PCSE methodology outlined above, and offers a different way to construct an estimate of the system covariance matrix, Ω, so potentially accommodating more general correlation structures.

---

[7] Maximum likelihood estimation can suffer from problems of non-convergence and is often sensitive to model misspecification, for example. The block bootstrap is sensitive to the size of the re-sampling block chosen, and also suffers from a ‘join-point’ problem, see Andrews (2002), which combined reduces the effectiveness of the block bootstrap compared to the improvement obtainable with non-dependent data using a parametric bootstrap.

3 Vector Moving Average Errors in a SoE

Consider the errors of a system of equations, as in (2), following a vector moving average (VMA) process of order Q, with heteroskedasticity of unspecified form:

$$u_{\cdot t} = I_n \varepsilon_{\cdot t} + \Theta_1 \varepsilon_{\cdot t-1} + \Theta_2 \varepsilon_{\cdot t-2} + \cdots + \Theta_Q \varepsilon_{\cdot t-Q} \tag{13}$$

with contemporaneous composite error vector:

$$\underset{n\times 1}{u_{\cdot t}} = \left(u_{1t}, u_{2t}, \ldots, u_{nt}\right)' \tag{14}$$

contemporaneous disturbance vectors:

$$\underset{n\times 1}{\varepsilon_{\cdot t}} = \left(\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt}\right)' \tag{15}$$

distributed as:

$$E\left(\varepsilon_{\cdot t}\varepsilon'_{\cdot t}\right) \sim N\left(0, \Sigma_t\right) \tag{16}$$

with associated VMA coefficient matrices[8]:

$$\Theta_q = \begin{bmatrix} \theta^q_{11} & \theta^q_{12} & \cdots & \theta^q_{1n} \\ \theta^q_{21} & \theta^q_{22} & \cdots & \theta^q_{2n} \\ \vdots & & & \vdots \\ \theta^q_{n1} & \cdots & \cdots & \theta^q_{nn} \end{bmatrix} \text{ for } q = \{1, 2, \ldots, Q\} \tag{17}$$

and contemporaneous (time-dependent) variance-covariance matrix:

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} & \cdots & \sigma_{1n,t} \\ \sigma_{21,t} & \sigma_{22,t} & \cdots & \sigma_{2n,t} \\ \vdots & & & \vdots \\ \sigma_{n1,t} & \sigma_{n2,t} & \cdots & \sigma_{nn,t} \end{bmatrix} \tag{18}$$

The VMA(Q) process, (13), is assumed invertible, with all roots lying outside the unit circle, so guaranteeing a unique process for a given covariance generating function.

---

[8] A standard MA(Q) process in each equation restricts the coefficient matrices Θ_q, (17), to be diagonal.

The (heteroskedastic) generally non-symmetric[9] covariance matrix at lag p of the composite error can be written:

$$\begin{aligned}\Gamma_{p,t} &= E\left(u_{\cdot t}u'_{\cdot t-p}\right)\\ &= \begin{cases}\sum_{k=0}^{Q-p}\Theta_{k+p}\Sigma_{t-p-k}\Theta'_{k} \text{ where } \Theta_0 = I_n \text{ for } p \le Q\\ 0 \text{ for } p > Q\end{cases}\end{aligned} \tag{19}$$

Assuming non-stochastic, fixed, regressors for ease of exposition in all that follows, the system covariance matrix Ω defined in (5) is linked to these lag-specific covariance matrices as follows:

$$\begin{aligned}X'E\left(uu'\right)X &= \sum_{p=1}^{Q}\sum_{t=1}^{T}\mathbf{x}'_{\mathbf{t}}E\left(u_{\cdot t}u'_{\cdot t-p}\right)\mathbf{x}_{\mathbf{t-p}}\\ X'\Omega X &= \sum_{p=1}^{Q}\sum_{t=1}^{T}\mathbf{x}'_{\mathbf{t}}\Gamma_{p,t}\mathbf{x}_{\mathbf{t-p}}\end{aligned} \tag{20}$$

The problem of accommodating a VMA error process can be reduced to consistently estimating $X'\Omega X/T$, so side-stepping the need for fully parametric inference involving the estimation of the VMA coefficients in (17)[10]. The following sub-section discusses how this is currently achieved in the case of robust inference for a single-equation while sub-section 3.2 shows how this can be extended to the case of system robust inference.

---

[9] For a VMA(1) process,

$$\Gamma_{1,t} = \begin{bmatrix} E\left(u_{it}u_{i,t-1}\right) & E\left(u_{it}u_{j,t-1}\right)\\ E\left(u_{it}u_{j,t+1}\right) & E\left(u_{jt}u_{j,t-1}\right)\end{bmatrix} = \begin{bmatrix}\theta^1_{11}\sigma_{11,t}+\theta^1_{12}\sigma_{21,t} & \theta^1_{11}\sigma_{12,t}+\theta^1_{12}\sigma_{22,t}\\ \theta^1_{21}\sigma_{11,t}+\theta^1_{22}\sigma_{21,t} & \theta^1_{21}\sigma_{12,t}+\theta^1_{22}\sigma_{22,t}\end{bmatrix},$$ which is clearly non-symmetric unless past shocks have an equivalent effect to future shocks.

[10] See Nicholls *et al* (1975) for a review of appropriate methods.

3.1 Single Equation HAC

For the specific case of a single equation, Newey and West (1987)[11] propose a heteroskedasticity and autocorrelation asymptotically robust covariance matrix for the regression parameters which can be calculated as[12]:

$$TVar\left(\hat{\boldsymbol{\beta}}_{\mathbf{i}}\right) = T\left(X_i'X_i\right)^{-1} X_i'\hat{\Omega}_i X_i \left(X_i'X_i\right)^{-1} \tag{21}$$

where:

$$T^{-1}\left(X_i'\hat{\Omega}_i X_i\right) = T^{-1}\left(\hat{\Upsilon}_{i0} + \sum_{p=1}^{T-1} w_{ip}\left[\hat{\Upsilon}_{ip} + \hat{\Upsilon}_{ip}'\right]\right) = \hat{S}_i \tag{22}$$

and,

$$\hat{\Upsilon}_{ip} = \sum_{t=p+1}^{T} \mathbf{x}_{\mathbf{i,t}}'\hat{u}_{i,t}\hat{u}_{i,t-p}\mathbf{x}_{\mathbf{i,t-p}} \tag{23}$$

where $\hat{u}_{it}$ are residuals from equation-by-equation OLS, $\mathbf{x}_{\mathbf{it}}$ is a typical $\left(1 \times k_i\right)$ row vector from the X_i matrix, w_{ip} is the scalar kernel (weighting function). Note that (21) is the estimated OLS coefficient variance formula, which as discussed above, is also employed in the PCSE methodology with either the SUR or Parks type estimate of the system-covariance matrix, Ω.

The kernel has the general form:

$$w_{ip} = f\left({}^{p}\!/_{m_i}\right) \tag{24}$$

where m_i is a ‘lag-truncation’ or bandwidth parameter. This parameter can be thought of as approximating the order of the (V)MA process[13], with $w_{ip} = 0,\ p > m_i$ for all commonly-used kernels except the Quadratic-Spectral kernel[14]. Newey and West (1987) suggest choosing a bandwidth as one plus the order of the process.

Bandwidth estimation methods, such as that proposed by Andrews (1991), can also be used to approximate this parameter. Such methods have the advantage that the bandwidth is allowed to grow at an appropriate rate with the sample size. Andrews (1991) states that “good

---

[11] Combining the work of Hansen and Hodrick (1980) and White (1980)

[12] Introducing an equation subscript to emphasise estimation of single-equation robust variances.

[13] Assuming the order of the process is known, Newey and West (1987) use $m = Q + 1$.

[14] Other popular kernel choices include the Parzen, Tukey-Hanning, Bartlett and Truncated kernel. See Davidson (2001) for the corresponding formulas.

performance of a HAC estimator […] only requires the automatic bandwidth parameter to be near the optimal bandwidth value and not precisely equal to it". If the optimal bandwidth equates, or is close to, the order of the (V)MA process, it would be equally appropriate to use this.

An appropriate kernel is essential for guaranteeing that the resulting covariance matrix of the regression parameters is positive semi-definite[15]. As $p \to \infty,\ w_{ip} \to 0$, ensuring that distant pseudo-covariances receive little or no weight compared to close-to pseudo-covariances. It can also be used to smooth this decline. In effect, for all but a Truncated kernel, this gives the sequence of residual-cross products the properties of a mixing-sequence.

The covariance matrix form calculated by (22) is consistent, as proved by Newey and West (1987), given suitable regularity conditions summarised in Hansen (1992). That is,

$$T^{-1}\left(X_i'\Omega_i X_i\right) - \hat{S}_i \xrightarrow{p} 0 \qquad (25)$$

Application of the Newey and West equation-by-equation methodology will produce heteroskedasticity and autocorrelation robust inference for a system of equations **only** where the regressor matrix is block-diagonal, as in (3), implying estimated coefficients measure *within* and not *between* equation effects[16] or if equations are uncorrelated. Furthermore, as this method does not produce a robust system covariance matrix, robust post-estimation tests of cross-equation restrictions are not possible.

It is commonplace in survey and cross-sectional time-series analysis to test whether a regression coefficient (an effect) in one equation is significantly different from that in another equation at the time of estimation, perhaps reparametrising the model using dummy-variables. Inference in such situations implicitly involves accounting for *between* (not just *within*) equation correlations. The SUR, Parks and PCSE methods, discussed in the previous section, will accommodate certain *within* and *between* equation correlation structures though,

---

[15] The Truncated (flat) kernel does not necessarily produce a positive-semi-definite covariance matrix estimate. Lin *et al* (2005) have recently suggested a modified version which may address this shortcoming though this is not considered further. Andrews (1991) notes that the Truncated kernel is theoretically superior to all other kernels when it produces a positive-semi-definite covariance matrix, so the work of Lin *et al* (2005) is worth consideration for future research.

[16] Whether the VMA coefficient matrices, Θ_q, are diagonal or not does not change this result. Likewise, if the regressor matrix is diagonal and the VMA coefficient matrices are not diagonal the Newey and West methodology applied to each equation will produce identical equation inference to system-HAC.

as highlighted, fall short of total data-dependent asymptotically robust inference. The Newey and West (1987) method, discussed above, does not account for *between* equation correlations, and so is not applicable in situations where the regressor matrix is not block-diagonal, which arises when dummy variables are used to capture cross-equation differences.

Extending (22) to allow general applicability regardless of system (and residual correlation) structure is not straightforward since, as demonstrated by (23), the methodology uses same equation residual and regressor cross-product terms to estimate a consistent covariance matrix $X_i \hat{\Omega}_i X_i$ for that equation. For general applicability, it is necessary to separate the regressors from the residuals, and identify the implicit correction to the system-error terms covariance matrix, Ω, which will have general form (5). The following sub-section shows how robust system inference can be achieved.

3.2 System-HAC

As a starting point, for our system approach, note that the White (1980) heteroskedasticity robust covariance matrix, in effect, replaces unknown disturbances by squared residuals[17]. Therefore, for a system-of-equations, consider

$$\underset{T\times T}{\hat{\Omega}_{ij}} = \begin{bmatrix} \hat{u}_{i,1}\hat{u}_{j,1} & 0 & \cdots & 0 \\ 0 & \hat{u}_{i,2}\hat{u}_{j,2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \hat{u}_{i,T}\hat{u}_{j,T} \end{bmatrix} \tag{26}$$

where (26) is a typical block of the heteroskedasticity robust system covariance matrix,

$$\hat{\Omega} = \left\{\hat{\Omega}_{ij}\right\} = \left\{\hat{\sigma}_{ij}\hat{M}_{ij}\right\} \tag{27}$$

Note, however, (26) does not account for autocorrelation. Applying the Newey and West (1987) methodology to $\hat{\Omega}_{ij}$, using as an example a bandwidth of 2 and kernel which truncates pseudo-covariances beyond the bandwidth, produces,

$$\hat{\Omega}_{ij} = \begin{bmatrix} \hat{u}_{i,1}\hat{u}_{j,1} & w_1\left(\hat{u}_{i,1}\hat{u}_{j,2}\right) & w_2\left(\hat{u}_{i,1}\hat{u}_{j,3}\right) & 0 & \cdots & 0 \\ w_1\left(\hat{u}_{j,2}\hat{u}_{i,1}\right) & \hat{u}_{i,2}\hat{u}_{j,2} & w_1\left(\hat{u}_{i,2}\hat{u}_{j,3}\right) & w_2\left(\hat{u}_{i,2}\hat{u}_{j,4}\right) & \ddots & \vdots \\ w_2\left(\hat{u}_{j,3}\hat{u}_{i,1}\right) & w_1\left(\hat{u}_{j,3}\hat{u}_{i,2}\right) & \hat{u}_{i,3}\hat{u}_{j,3} & w_1\left(\hat{u}_{i,3}\hat{u}_{j,4}\right) & \ddots & 0 \\ 0 & w_2\left(\hat{u}_{j,4}\hat{u}_{i,2}\right) & w_1\left(\hat{u}_{j,4}\hat{u}_{i,3}\right) & \hat{u}_{i,4}\hat{u}_{j,4} & \ddots & w_2\left(\hat{u}_{i,T-2}\hat{u}_{j,T}\right) \\ \vdots & 0 & \vdots & \vdots & \ddots & w_1\left(\hat{u}_{i,T-1}\hat{u}_{j,T}\right) \\ 0 & \cdots & 0 & 0 & \cdots & \hat{u}_{i,T}\hat{u}_{j,T} \end{bmatrix} \tag{28}$$

or more generally:

$$\left\{\hat{\Omega}_{ij,ts}\right\} = \begin{cases} \hat{u}_{it}\hat{u}_{js} \text{ for } t = s \\ w_p\hat{u}_{it}\hat{u}_{js} \text{ for } t \neq s \text{ where } p = \left|s-t\right| \end{cases} \tag{29}$$

where, as before, $\hat{u}_{it}$ are the residuals from equation-by-equation OLS regressions. The weight function, does not have an equation subscript since this is common across all equations to ensure that this system covariance matrix is positive semi-definite (see den Haan and Levin 1997).

Although consistency is not formally established here, it is assumed to result from similar conditions to those required for consistency in the single-equation robust covariance matrix context. (publication add this proof and proof of positive semi-definiteness?)

---

[17] Davidson and MacKinnon (1985) suggest improvements to finite sample performance of this heteroskedasticity robust covariance matrix may be achieved by appropriate scaling of the squared-residual terms. This refinement is not considered further.

As discussed in the single equation context, the bandwidth parameter can be approximated by one-plus the order of the VMA process, if this is known. Alternatively, the bandwidth can be chosen by selecting the largest bandwidth in the set of bandwidths calculated for each equation in the system, using a (single-equation) ‘automatic’ data-dependent bandwidth methodology, such as that proposed by Andrews (1991)[18].

In forming a robust system covariance matrix the method outlined above has the further advantage that consistent post-estimation tests of cross-equation restrictions can be conducted using the standard robust version of the Wald test (see Wooldridge 2002a or Godfrey and Orme 2003).

Removing some degree of the residual correlation prior to estimating the asymptotically robust covariance matrix *may* improve finite sample performance. Such methods include the prewhitening methodology of Andrews and Monhan (1992) and the VARHAC methodology of den Haan and Levin (2000), which in effect differ only in the process used to estimate the asymptotically robust covariance matrix, the former using kernel based methods, as above, the latter, not. These methods will produce consistent inference for a system-of-equations only where the regressor matrix is block-diagonal. Despite the shortcomings of such methodologies, as discussed by Sul, Phillips and Choi (2005) and Newey and West (1994), further research would be useful to extend their applicability to the situation of a non-block-diagonal regressor matrix to allow for a full comparison of the effects of initially filtering the residuals to remove some inter-dependency.

In conclusion, asymptotic heteroskedasticity, *within* and *between* equation robust inference can be achieved for a system-of-equations regardless of the structure of the regressor matrix using (29). In the single-equation case, this is equivalent to the Newey and West (1987) methodology. Using a data-dependent bandwidth methodology for bandwidth choice, such as that proposed by Andrews (1991) and modified for applicability to the system-context, researchers can achieve robust inference without having to assume a specific correlation or heteroskedasticity structure.

---

[18] In practice, parameters from an AR(1) regression involving the estimated residuals are used in the formula provided by Andrews (1991) to calculate this bandwidth. In the VMA setting, this would require finding the optimum bandwidth for each equation and choosing the maximum. Although Andrews suggests other structural forms for calculating the bandwidth, it is the AR(1) form which tends to be utilised.

The following section provides Monte Carlo simulation evidence for the size and power properties of system heteroskedasticity and autocorrelation robust inference, extending the single equation study of Smith and Yadav (1996).

4 System-HAC Performance Evaluation

Using Monte Carlo simulation techniques, the performance of the system heteroskedasticity and autocorrelation (HAC) robust methodology, outlined in the previous section, is evaluated.

Building on the single equation simulation results of Smith and Yadav (1996), this section explores the finite-sample performance for a range of data-generating processes (DGPs). Potentially robust methods are compared in terms of test size and size adjusted power. The performance of the Bartlett kernel function, advocated by Newey and West (1987), is compared with the performance of the theoretically superior[19] Quadratic-Spectral kernel function, recommended by Andrews (1991), along with two other commonly used kernels. The DGPs analysed are chosen mindful of the evidence of Galbraith and Zinde-Walsh (1994) that closeness to non-invertibility can adversely affect approximations of a potentially robust covariance-matrix estimator. Performance under three types of disturbance heteroskedasticity is also examined.

Experiments are designed as follows. A two-equation system, $i=\{1,2\}$ is constructed, with each equation containing a constant term and an equation-specific regressor:

$$\begin{aligned} y_{1t} &= \alpha_1 + x_{1t}'\beta_1 + u_{1t} \\ y_{2t} &= \alpha_2 + x_{2t}'\beta_2 + u_{2t} \end{aligned} \tag{30}$$

stacked to form a system as in (2) where

$$\underset{2T\times 1}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad \underset{2T\times 4}{X} = \begin{bmatrix} X_1 & 0 \\ X_2 & X_2 \end{bmatrix} \quad \underset{4\times 1}{\beta} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \quad \underset{2T\times 1}{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{31}$$

and X_i is the $(T\times 2)$ matrix of equation specific regressors. Through the specific (non-block-diagonal) construction of the regressor matrix, the estimated coefficients for the second equation will measure the difference from those obtained in the first equation, so that

$$\begin{aligned} \gamma_1 &= [\alpha_1 \quad \beta_1]' \\ \gamma_2 &= [\alpha_2 - \alpha_1 \quad \beta_2 - \beta_1]' = [\tau_1 \quad \tau_2]' \end{aligned} \tag{32}$$

Measuring cross-equation differentials in this way is relevant where the regressors capture an identical feature across equations with equation specific effects (eg. country GDP). It should be noted that equation-by-equation OLS coefficient estimates are unbiased in the presence of heteroskedasticity and autocorrelation, assuming non-stochastic regressors.

---

[19] In the class of kernel functions guaranteeing positive semi-definite covariance matrices.

Coefficient values are as follows:

$$\begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \text{ therefore } \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{33}$$

where all hypothesis tests are conducted against a two-tailed alternative of non-equality with the true parameter value and are performed at the 5% level of significance.

Consider that the system errors following a VMA(3) process, as described by equations (13) to (18) in the previous section. The initially homoskedastic contemporaneous VCV matrix of the system disturbances is given the following specification:

$$\Sigma = E\left(\varepsilon_t \varepsilon_t'\right) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.9 \\ 0.9 & 1.5 \end{bmatrix} \tag{34}$$

Note that equation 2 has a larger variance than equation 1, and the contemporaneous correlation between equations is relatively strong ($\rho_{1,2} = 0.735$).

The simulations employ 8 combinations of VMA error coefficients, as outlined in Table 4.1, with associated characteristic equation roots shown in Table 4.2. The first six specifications have a diagonal VMA coefficient matrix (17) for each lag, which is equivalent to a MA process in each (contemporaneously correlated) equation. The final two specifications are full VMA(3) processes (with non-diagonal coefficient matrices). A process with spherical errors is included to quantify inferiority versus OLS in this situation.

As shown by the characteristic equation roots in Table 4.2, the coefficients are chosen to provide a range of processes with roots close to, far away, or on the unit circle (and so non-invertible).

Table 4.1: (V)MA Error Process Parameters

Equation Reference	VMA Coefficients											
	Θ_1				Θ_2				Θ_3			
	θ_{11}^1	θ_{12}^1	θ_{21}^1	θ_{22}^1	θ_{11}^2	θ_{12}^2	θ_{21}^2	θ_{22}^2	θ_{11}^3	θ_{12}^3	θ_{21}^3	θ_{22}^3
MAe10	1	0	0	1	1	0	0	1	1	0	0	1
MAe9	2.7	0	0	2.7	2.43	0	0	2.43	0.729	0	0	0.729
MAe5	1.5	0	0	1.5	0.75	0	0	0.75	0.125	0	0	0.125
MAe951	1.5	0	0	1.5	0.59	0	0	0.59	0.045	0	0	0.045
MAe1	0.3	0	0	0.3	0.03	0	0	0.03	0.001	0	0	0.001
NoMA	0	0	0	0	0	0	0	0	0	0	0	0
VMAe1	1.6	0.2	0.4	0.4	0.7	0.1	0.2	0.2	0.05	0.05	0.01	0.05
VMAe2	1.9	1	1	0.6	0.4	0.5	0.6	0.1	-0.2	0.25	-0.1	-0.02

Table 4.2: Characteristic Equation Roots

Equation Reference	Roots					
	z_1	z_2	z_3	z_4	z_5	z_6
MAe10	1	1	1	1	1	1
MAe9	0.9	0.9	0.9	0.9	0.9	0.9
MAe5	0.5	0.5	0.5	0.5	0.5	0.5
MAe951	0.9	0.9	0.5	0.5	0.1	0.1
MAe1	0.1	0.1	0.1	0.1	0.1	0.1
NoMA	0	0	0	0	0	0
VMAe1	0.11	0.22	0.58	0.98	0.38*	0.38*
VMAe2	0.69	0.49#	0.68	0.13	0.48*	0.48*

Notes: all roots are absolute values. * denotes the absolute value of a complex root, and # denotes the invertible characteristic equation counterpart root (ie. the reciprocal of the root identified).

Non-constant regressors are generated from an equation-specific AR(1) process[20]:

$$x_{it} = \phi \, x_{it-1} + \zeta_{it} \qquad (35)$$

$\xi_{it} \sim NID(0,1)$, $x_{i0} = 0$. The autoregressive coefficient takes the following values (for both equations): $\phi = \{0.25, 0.50\}$. Observations are generated[21] for the two sample sizes: $T = \{100, 300\}$.

---

[20] If regressors are serially-uncorrelated cross-equation restrictions on the slope-coefficient only can be consistently tested without accounting for the VMA error structure (test power, is affected by the VMA structure). This correct sizing is indicated by the analysis of Smith and Yadav (1996) and additional simulations conducted for the present research (results not shown). This is not the case for tests on the intercept coefficient.

[21] To reduce the effect of the initial observation, which is generated as a random draw from a standard Normal distribution, an additional 100 observations are generated with the first 100 then being discarded.

Performance under the following three forms of heteroskedasticity is considered:

$$\begin{aligned} &\text{No HET:} \quad E\left(\varepsilon_i^2\right) = \sigma_{ii} \\ &\text{HET(1):} \quad E\left(\varepsilon_i^2\right) = \begin{cases} \sigma_{ii} \text{ for } i = 1,\dots,T/2 \\ 4\sigma_{ii} \text{ for } i = \left(T/2\right)+1,\dots,T \end{cases} \\ &\text{HET(2):} \quad E\left(\varepsilon_i^2\right) = \sigma_{ii} x_i^2 \end{aligned} \tag{36}$$

Note that type NoHET is equivalent to homoskedastic disturbances, and HET(1) type heteroskedasticity is simply a variance break. HET(2) type, since from (35) the regressors are autocorrelated, is effectively inducing an auto-regressive conditional heteroskedasticity, ARCH(1) type structure, to the disturbance variances. 2,000 replications are performed for each experiment.

To implement system-HAC, the bandwidth, m_i in (24), can be chosen as the maximum of the equation-by-equation bandwidths calculated using the automatic bandwidth method of Andrews (1991). As shown in the single-equation context by Smith and Yadav (1996), this method produces similar finite sample test slope coefficient test size-performance compared to using the fixed ‘one plus MA order’ bandwidth methodology, advocated by Newey and West (1987) but has the benefit of being applicable where the order of the process in unknown. Test size for both the slope and intercept coefficients will be compared using both bandwidth methodologies and a range of popular kernels functions: Bartlett, Quadratic-Spectral (QS), Tukey-Hanning (TH) and Parzen.

Methods compared to system-HAC are PCSE/SUR (which has no autocorrelation correction) and PCSE with an AR(1) error correction, PCSE/Parks (see (12) with (8) and (9) for the construction of $\hat{\Omega}$, respectively).

This performance evaluation section proceeds as follows: section 4.1 analyses test size, section 4.2 analyses test power, and section 4.2.1 summarises the finding from these results.

4.1 Size

This section examines test size for a sample of 300 observations in sub-section 4.1.1 and for 100 observations in sub-section 4.1.2. For both sample-sizes, current system methods, outlined previously, are compared with the system-HAC methodology employing the Andrews (1991) automatic bandwidth methodology and a quadratic-spectral kernel. As we are particularly interested in the performance of system-HAC with full VMA processes (VMAe1 and VMAe2 as detailed in Table 4.1) simulations results are obtained for the full VMA processes with all three types of heteroskedasticity outlined in (36). Other processes are compared in the context of no-heteroskedasticity. Finally, for both sample sizes, system-HAC bandwidth calculation method and kernel choice are compared.

4.1.1 Sample Size of 300 Observations

For all simulations the nominal test size is set at 5%. Table 4.3 presents the result for tests of cross-equation restrictions using a large sample (300 observations) with a range of DGPs as discussed above. From (32), it should be noted that testing the significance of τ_1 is equivalent to testing whether there is a significant difference in the intercept terms between the two equations, while the significance of the difference in the slope coefficients is tested through τ_2.

Allowing for variability around the nominal size level due to the simulation process, Table 4.3 demonstrates how for both the intercept and slope differential coefficient, the system-HAC tests are nearly always correctly sized regardless of the DGP in a sample of $T = 300$. Minor oversizing for the system-HAC methodology is observed with complex heteroskedasticity: HET(2).

Table 4.3: Large Sample Empirical Sizes of Cross-Equation Regressor Significance Tests

	Regressor AR Coef.	**VMA Process Ref.**	**PCSE/SUR**		**PCSE/Parks**		**HAC (QS)**	
			τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
No-HET	0.25	NoMA	**0.054**	**0.050**	**0.065**	**0.060**	**0.058**	**0.053**
	0.25	MAe1	0.072	**0.065**	**0.044**	**0.049**	**0.057**	**0.056**
	0.25	MAe9	0.125	0.083	0.107	0.022	**0.069**	**0.057**
	0.25	MAe10	0.139	0.105	0.110	**0.040**	**0.060**	**0.070**
	0.25	VMAe1	0.097	0.089	0.004	**0.065**	**0.065**	**0.063**
	0.25	VMAe2	0.080	0.081	0.008	0.072	**0.061**	**0.058**
	0.5	NoMA	**0.051**	**0.049**	**0.059**	**0.059**	**0.051**	**0.051**
	0.5	MAe1	0.077	0.080	**0.049**	**0.054**	**0.052**	**0.058**
	0.5	MAe5	0.141	0.146	0.072	**0.040**	**0.066**	**0.067**
	0.5	MAe9	0.158	0.143	0.096	**0.044**	**0.064**	**0.061**
	0.5	MAe10	0.185	0.170	0.093	**0.047**	**0.069**	**0.069**
	0.5	MAe951	0.131	0.138	**0.057**	**0.042**	**0.066**	**0.064**
	0.5	VMAe1	0.132	0.130	0.013	**0.054**	**0.065**	**0.068**
	0.5	VMAe2	0.118	0.113	0.018	**0.051**	**0.065**	**0.067**
HET(1)	0.25	VMAe1	0.095	0.087	0.003	**0.064**	**0.059**	**0.061**
	0.25	VMAe2	0.082	0.079	0.009	**0.066**	**0.058**	**0.057**
	0.5	VMAe1	0.127	0.131	0.009	**0.057**	**0.059**	**0.059**
	0.5	VMAe2	0.117	0.119	0.019	**0.054**	**0.060**	**0.059**
HET(2)	0.25	VMAe1	**0.043**	0.112	**0.044**	**0.057**	**0.063**	0.071
	0.25	VMAe2	**0.049**	0.097	**0.051**	**0.062**	**0.059**	**0.058**
	0.5	VMAe1	**0.065**	0.168	0.023	**0.054**	**0.064**	0.075
	0.5	VMAe2	**0.055**	0.150	0.028	**0.056**	**0.065**	0.075

Notes: Sample size of 300 observations with 2,000 replications. Empirical sizes in the range 3% to 7% are shown in bold. All results refer to testing the significance of τ_1 and τ_2 in (31) with respect to the true values set. HAC results obtained using a modified Andrews (1991) 'automatic' bandwidth calculation. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. Heteroskedasticity forms relate to equation (36). Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

Approximating the VMA process by an AR(1) process and using the Parks correction methodology, PCSE/Parks, results in tests on the intercept differential which are almost always incorrectly sized, and are strongly undersized for the full VMA processes (VMAe1, VMAe2). Generally, tests on the slope coefficient differential are close to the nominal size for this method when there is stronger (0.5) regressor autocorrelation, suggesting that this correlation might dominate the effects of the moving-average error correlations.

For a spherical error process, regardless of the strength of the AR process generating the regressor, the PCSE/SUR method and the system-HAC method both result in tests with approximately correct sizing. This suggests that in reasonably large samples, there is no distinguishable impact on test size in applying a HAC covariance matrix even if the presence of residual correlation is only doubtful.

It is interesting to note from Table 4.3 that intercept differential test sizes using the PCSE/SUR methodology are close to the asymptotic, nominal, test size for VMA processes with complex, HET(2) type, heteroskedasticity. It remains to be understood why this is the case, and whether it is not just a quirk of the particular simulation methodology employed. Apart from this anomaly, PCSE/SUR are oversized for all except the non-VMA (NoMA) process.

Table 4.4 presents simulation results for tests of significance of α_1 and β_1, again involving a relatively large sample of 300 observations. Unlike Table 4.3 which presented results for τ_1 and τ_2 which were implicitly tests of cross-equation restrictions, Table 4.4 results refer to within-equation restrictions. Tests on these coefficients will not require the use of cross-equation covariance terms, and so it is expected that test sizes should be similar to those observed using single-equation HAC methods. Although Smith and Yadav (1996) provide simulation evidence regarding the impact of methodological differences on test size in the single-equation context for the specific case an equation of type MAe10, our analysis extends their work to examine the intercept test size, in a system context. The following section further extends this analysis by investigating size-adjusted power.

Table 4.4 again demonstrates that, for the wide range of DGPs examined, the system-HAC method generally yields tests with good size properties. In a small number of cases these tests are oversized, though compared to other methodologies the variability in sizing is relatively small. Comparing system-HAC test size performance to the single-equation HAC test size results obtained by Smith and Yadav (1996), it appears that these methods have similar test size performance.

As previously noted in the cross-equation context, heteroskedasticity unless complex HET(2) type, does not clearly influence the performance of system-HAC slope test size. The anomaly of non-autocorrelation corrected PCSE/SUR methodology intercept test size being close to the nominal test size in this case is again observed.

Table 4.4: Large Sample Empirical Size of Within-Equation Regressor Significance Tests

	Regressor AR Coef.	**VMA Process Ref.**	**PCSE/SUR**		**PCSE/Parks**		**System-HAC (QS)**	
			α_1	β_1	α_1	β_1	α_1	β_1
No-HET	0.25	NoMA	**0.046**	**0.039**	**0.051**	**0.048**	**0.050**	**0.044**
	0.25	MAe1	0.082	**0.069**	**0.052**	**0.055**	**0.055**	**0.065**
	0.25	MAe9	0.167	0.103	0.100	0.022	**0.057**	**0.064**
	0.25	MAe10	0.216	0.109	0.113	**0.040**	**0.066**	**0.069**
	0.25	VMAe1	0.151	0.094	0.019	**0.065**	**0.065**	**0.066**
	0.25	VMAe2	0.128	0.082	**0.068**	**0.070**	**0.066**	**0.063**
	0.5	NoMA	**0.044**	**0.044**	**0.051**	**0.039**	**0.051**	**0.050**
	0.5	MAe1	0.081	0.074	**0.049**	**0.058**	**0.058**	**0.063**
	0.5	MAe5	0.177	0.148	0.071	**0.041**	**0.066**	0.072
	0.5	MAe9	0.182	0.154	0.097	**0.035**	0.071	0.072
	0.5	MAe10	0.220	0.165	0.107	**0.049**	**0.070**	0.077
	0.5	MAe951	0.162	0.142	**0.056**	**0.044**	**0.063**	**0.069**
	0.5	VMAe1	0.168	0.139	0.020	**0.059**	0.074	0.073
	0.5	VMAe2	0.144	0.121	**0.057**	**0.056**	0.071	**0.070**
HET(1)	0.25	VMAe1	0.152	0.092	**0.030**	**0.068**	**0.060**	**0.059**
	0.25	VMAe2	0.127	0.084	0.096	0.071	**0.060**	**0.059**
	0.5	VMAe1	0.162	0.141	**0.037**	**0.061**	**0.068**	**0.070**
	0.5	VMAe2	0.137	0.121	0.095	**0.053**	**0.070**	**0.064**
HET(2)	0.25	VMAe1	**0.046**	0.114	**0.061**	**0.061**	**0.067**	0.074
	0.25	VMAe2	**0.043**	0.104	**0.056**	**0.061**	**0.059**	0.071
	0.5	VMAe1	**0.056**	0.175	0.029	**0.051**	**0.060**	0.080
	0.5	VMAe2	**0.039**	0.156	0.025	**0.055**	**0.054**	0.071

Notes: Sample size of 300 observations with 2,000 replications. Empirical sizes in the range 3% to 7% are shown in bold. All results refer to testing the significance of α_1 and β_1 in (31) with respect to the true values set. HAC results obtained using a modified Andrews (1991) ‘automatic’ bandwidth calculation. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. Heteroskedasticity forms relate to equation (36). Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

In conclusion, for system-HAC, results in Table 4.4 follow a similar pattern to those documented for cross-equation restrictions in Table 4.3, which suggests that the correction to the system covariance matrix implicit in the single equation context is not fundamentally different in terms of test performance from the correction applied to the cross-equation covariance terms. Empirical size properties for the PCSE/Parks method do differ, however, particularly for full VMA processes (VMAe1, VMAe2) and those with HET(1) type heteroskedasticity, with serious undersizing apparent for tests on the intercept coefficient for these processes between equations but not always for within equation tests.

Table 4.5 and Table 4.6 compare, for a reasonably large sample of 300 observations, the modified data-dependent Andrews (1991) 'automatic' bandwidth methodology with the 'manual' methodology of Newey and West (1987) which employs the order of the (V)MA process. This comparison is pertinent as in cases where the order of the (V)MA process is unknown it is useful to quantify any cost to approximating the bandwidth from the data. These tables also compare a range of popular kernels.

Results suggest that there is little difference in the performance of either the various kernels or the bandwidth methodology employed for the range of DGPs tested when a reasonably large sample is available. This may result from the order of the process being small relative to the sample size, which may equate the benefits associated to the controlled growth of this bandwidth parameter from the automatic versus the manual method. In terms of kernel used, the Quadratic-Spectral (QS) kernel advocated by Andrews (1991) does produce slightly more accurately sized tests, though this difference tends to be small and on the whole would lead to the conclusion that kernel choice does not markedly influence test performance.

Comparing Table 4.5 for between-equation tests with Table 4.6 for within-equation tests again indicates that system-HAC properties are near identical regardless of whether tests involve the cross-equation covariance terms. Again, this probably suggests that the system-HAC and single-equation HAC do not have fundamentally different properties, since the within-equation results will be equivalent to those obtained using single-equation HAC.

Table 4.5: Large Sample Empirical Size of Between-Equation Regressor Significance Tests for Various Kernels and Bandwidth Methods

	Regressor AR Coef.	VMA Process Ref.	Newey & West (1987) Manual Bandwidth								Andrews (1991) Automatic Bandwidth							
			HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)		HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)	
			τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
No-HET	0.25	NoMA	**0.063**	**0.056**	**0.060**	**0.056**	**0.061**	**0.058**	**0.063**	**0.061**	**0.058**	**0.055**	**0.057**	**0.055**	**0.057**	**0.054**	**0.058**	**0.053**
	0.25	MAe1	**0.057**	**0.058**	**0.055**	**0.057**	**0.056**	**0.057**	**0.057**	**0.057**	**0.058**	**0.057**	**0.056**	**0.057**	**0.056**	**0.057**	**0.057**	**0.056**
	0.25	MAe9	0.072	**0.060**	0.075	**0.060**	**0.065**	**0.056**	**0.062**	**0.053**	**0.070**	**0.059**	**0.070**	**0.059**	**0.070**	**0.056**	**0.069**	**0.057**
	0.25	MAe10	0.073	0.072	0.080	0.072	0.071	**0.065**	**0.063**	**0.066**	**0.067**	**0.070**	**0.061**	**0.068**	**0.059**	**0.069**	**0.060**	**0.070**
	0.5	NoMA	**0.058**	**0.051**	**0.053**	**0.052**	**0.057**	**0.051**	**0.060**	**0.052**	**0.051**	**0.051**	**0.052**	**0.050**	**0.051**	**0.050**	**0.051**	**0.051**
	0.5	MAe1	**0.053**	**0.060**	**0.053**	**0.060**	**0.052**	**0.058**	**0.052**	**0.057**	**0.053**	**0.062**	**0.051**	**0.059**	**0.051**	**0.058**	**0.052**	**0.058**
	0.5	MAe9	0.075	**0.066**	**0.080**	**0.070**	**0.068**	**0.064**	**0.062**	**0.058**	**0.068**	**0.065**	**0.065**	**0.065**	**0.065**	**0.063**	**0.064**	**0.061**
	0.5	MAe10	0.086	0.086	0.092	0.095	0.081	0.079	**0.070**	**0.070**	0.074	0.072	0.072	**0.070**	**0.069**	**0.070**	**0.069**	**0.069**

Table 4.6: Large Sample Empirical Size of Within-Equation Regressor Significance Tests for Various Kernels and Bandwidth Methods

	Regressor AR Coef.	VMA Process Ref.	Newey & West (1987) Manual Bandwidth								Andrews (1991) Automatic Bandwidth							
			HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)		HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)	
			α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1
No-HET	0.25	NoMA	**0.051**	**0.047**	**0.048**	**0.047**	**0.050**	**0.048**	**0.054**	**0.051**	**0.050**	**0.045**	**0.049**	**0.045**	**0.049**	**0.044**	**0.050**	**0.044**
	0.25	MAe1	**0.059**	**0.066**	**0.059**	**0.066**	**0.055**	**0.065**	**0.052**	**0.063**	**0.059**	**0.065**	**0.054**	**0.064**	**0.055**	**0.064**	**0.055**	**0.065**
	0.25	MAe9	0.071	**0.071**	0.077	**0.070**	**0.064**	**0.065**	**0.058**	**0.059**	**0.063**	**0.067**	**0.058**	**0.067**	**0.056**	**0.065**	**0.057**	**0.064**
	0.25	MAe10	0.093	0.073	0.108	0.076	0.088	**0.070**	**0.076**	**0.071**	**0.074**	**0.074**	**0.067**	**0.072**	**0.067**	**0.070**	**0.066**	**0.069**
	0.5	NoMA	**0.053**	**0.051**	**0.054**	**0.051**	**0.053**	**0.052**	**0.056**	**0.052**	**0.051**	**0.050**	**0.051**	**0.050**	**0.051**	**0.050**	**0.051**	**0.050**
	0.5	MAe1	**0.061**	**0.063**	**0.061**	**0.064**	**0.058**	**0.062**	**0.058**	**0.063**	**0.059**	**0.065**	**0.057**	**0.064**	**0.057**	**0.063**	**0.058**	**0.063**
	0.5	MAe9	0.077	**0.085**	**0.087**	**0.091**	**0.071**	**0.077**	**0.067**	**0.071**	**0.075**	**0.075**	**0.074**	**0.073**	**0.072**	**0.071**	**0.071**	**0.072**
	0.5	MAe10	0.089	0.085	0.105	0.092	0.083	0.080	**0.070**	**0.074**	0.073	0.081	0.072	**0.079**	**0.067**	**0.076**	**0.070**	**0.077**

Notes (both tables): Sample size of 300 observations with 2,000 replications. ‘TH’ refers to the Tukey-Hanning kernel, and ‘QS’, the Quadratic-Spectral kernel. Empirical sizes in the range 3% to 7% are shown in bold. All results refer are tests of significance of the column headed coefficient from equation (31) with respect to the true values set. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. All within-equation error processes are homoskedastic. Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

4.1.2 Sample Size of 100 Observations

System-HAC is designed to produce an asymptotically consistent covariance matrix robust to autocorrelation and heteroskedasticity. It is to be expected that in finite samples system-HAC performance will deteriorate. Extending the analysis to investigate the performance of all methods in a smaller sample, Table 4.7 repeats the analysis of Table 4.3 for cross-equation restrictions with a sample of 100 observations.

Table 4.7: Small Sample Empirical Size of Cross-Equation Regressor Significance Tests

	Regressor AR Coef.	VMA Process Ref.	PCSE/SUR		PCSE/Parks		System-HAC (QS)	
			τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
No-HET	0.25	NoMA	**0.052**	**0.052**	**0.067**	**0.064**	**0.061**	**0.062**
	0.25	MAe1	0.075	**0.069**	**0.061**	**0.054**	0.073	0.075
	0.25	MAe9	0.147	0.102	0.156	0.025	0.094	0.090
	0.25	MAe10	0.134	0.094	0.160	**0.037**	0.088	0.089
	0.25	VMAe1	0.090	0.086	0.006	**0.057**	0.077	0.077
	0.25	VMAe2	0.071	0.079	0.009	**0.066**	0.072	0.075
	0.5	NoMA	**0.050**	**0.049**	**0.070**	**0.064**	**0.065**	**0.063**
	0.5	MAe1	0.086	0.093	**0.063**	**0.061**	0.082	0.087
	0.5	MAe5	0.165	0.158	0.130	**0.051**	0.098	0.112
	0.5	MAe9	0.178	0.166	0.133	**0.048**	0.112	0.110
	0.5	MAe10	0.170	0.164	0.137	**0.051**	0.108	0.109
	0.5	MAe951	0.158	0.148	0.102	**0.055**	0.099	0.111
	0.5	VMAe1	0.129	0.141	0.013	**0.064**	0.096	0.100
	0.5	VMAe2	0.114	0.124	0.022	**0.064**	0.090	0.093
HET(1)	0.25	VMAe1	0.088	0.080	0.007	**0.059**	0.072	0.076
	0.25	VMAe2	**0.069**	0.076	0.012	**0.065**	**0.070**	0.079
	0.5	VMAe1	0.128	0.141	0.017	**0.067**	0.095	0.097
	0.5	VMAe2	0.116	0.120	0.024	**0.061**	0.086	0.091
HET(2)	0.25	VMAe1	**0.042**	0.098	**0.047**	**0.060**	0.072	0.090
	0.25	VMAe2	**0.034**	0.076	**0.050**	**0.055**	**0.068**	0.081
	0.5	VMAe1	**0.059**	0.168	0.025	**0.057**	0.073	0.111
	0.5	VMAe2	**0.051**	0.137	0.029	**0.061**	0.072	0.088

Notes: Sample size of 300 observations with 2,000 replications. Empirical sizes in the range 3% to 7% are shown in bold. All results refer to testing the significance of τ_1 and τ_2 in (31) with respect to the true values set. HAC results obtained using a modified Andrews (1991) 'automatic' bandwidth calculation. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. Heteroskedasticity forms relate to equation (36). Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

For both intercept and slope coefficient differentials, finite sample performance of system-HAC as shown in Table 4.7 tends to be oversized compared to the nominal test size of 5%. Empirical sizes for the system-HAC methodology are up to twice the nominal level, and also appear to be increasing with respect to the AR regressor coefficient. This compares to the PCSE/Parks method, which for the slope differential tends to be correctly sized, but for the intercept differential suffers from extreme variability in size distortions, with notably severe undersizing with full VMA error processes and large oversizing for most MA processes. Again, the anomaly of complex, HET(2) type, heteroskedasticity yielding correct intercept differential test sizes for the non-autocorrelation corrected PCSE/SUR methodology is still evident.

Table 4.8 repeats the analysis for within-equation restrictions in a small sample context, for comparison with the performance evaluation for the large sample counterpart in Table 4.5.

Again Table 4.8 demonstrates performance of system-HAC is similar to that for the tests of cross-equation restrictions highlighting that the underlying correction of system-HAC is similar to that of single-equation HAC. The PCSE/Parks method again suffers from substantial oversizing for tests of the intercept coefficient, but the undersizing observed for VMA processes involving tests of cross-equation restrictions is largely nullified in the within equation test context. As before, PCSE/Parks is generally correctly sized for tests involving the slope coefficient.

Table 4.8: Small Sample Empirical Size of Within-Equation Regressor Significance Tests

	Regressor AR Coef.	VMA Process Ref.	PCSE/SUR		PCSE/Parks		System-HAC (QS)	
			α_1	β_1	α_1	β_1	α_1	β_1
No-HET	0.25	NoMA	**0.057**	**0.064**	**0.046**	**0.057**	**0.064**	**0.070**
	0.25	MAe1	0.092	**0.060**	**0.060**	**0.057**	0.084	**0.066**
	0.25	MAe9	0.176	0.100	0.128	0.026	0.098	0.089
	0.25	MAe10	0.218	0.095	0.162	**0.043**	0.102	0.082
	0.25	VMAe1	0.161	0.087	**0.042**	**0.063**	0.083	0.083
	0.25	VMAe2	0.125	0.077	0.096	0.071	0.080	0.081
	0.5	NoMA	**0.051**	**0.059**	**0.057**	**0.059**	**0.063**	**0.068**
	0.5	MAe1	0.088	0.083	**0.056**	**0.056**	0.084	0.076
	0.5	MAe5	0.185	0.154	0.107	**0.045**	0.111	0.106
	0.5	MAe9	0.193	0.158	0.121	**0.039**	0.111	0.108
	0.5	MAe10	0.226	0.165	0.155	**0.052**	0.113	0.102
	0.5	MAe951	0.172	0.141	0.086	**0.047**	0.109	0.099
	0.5	VMAe1	0.161	0.147	**0.036**	**0.057**	0.100	0.102
	0.5	VMAe2	0.136	0.117	0.085	**0.065**	0.097	0.092
HET(1)	0.25	VMAe1	0.157	0.087	0.071	**0.059**	0.086	0.083
	0.25	VMAe2	0.120	0.076	0.183	**0.069**	0.078	0.080
	0.5	VMAe1	0.155	0.144	**0.057**	**0.053**	0.100	0.102
	0.5	VMAe2	0.138	0.115	0.145	**0.054**	0.099	0.092
HET(2)	0.25	VMAe1	**0.053**	0.103	**0.053**	**0.053**	0.072	0.093
	0.5	VMAe2	**0.053**	0.181	**0.035**	**0.057**	**0.069**	0.115
	0.25	VMAe1	**0.040**	0.080	**0.050**	**0.048**	**0.065**	0.082
	0.5	VMAe2	**0.041**	0.156	0.029	**0.067**	**0.062**	0.090

Notes: Sample size of 300 observations with 2,000 replications. Empirical sizes in the range 3% to 7% are shown in bold. All results refer to testing the significance of α_1 and β_1 in (31) with respect to the true values set. HAC results obtained using a modified Andrews (1991) 'automatic' bandwidth calculation. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. Heteroskedasticity forms relate to equation (36). Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

Table 4.9 and

Table 4.10 investigate whether finite sample performance can be improved through kernel choice or bandwidth methodology, the equivalent results for a larger sample being shown in Table 4.5 and Table 4.6. Results suggest test size distortions are unrelated to kernel choice and hence this finding seems unrelated to sample size. There is also little perceivable difference attributable to bandwidth methodology which is useful in suggesting that the use of the true (V)MA process order, when this is known, will yield similar results to estimating a bandwidth designed to grow at an appropriate rate with the sample size. Further research would, however, be useful to uncover the point at which the automatic bandwidth is optimal, if, for example, the order of the (V)MA process is much larger compared to the sample size.

Table 4.9: Small Sample Between-Equation Regressor Significance Test Size Kernel and Bandwidth Calculation Comparison

	Regressor AR Coef.	VMA Process Ref.	Newey & West (1987) Manual Bandwidth								Andrews (1991) Automatic Bandwidth							
			HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)		HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)	
			τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
No-HET	0.25	NoMA	**0.069**	**0.068**	**0.068**	**0.065**	**0.069**	0.071	0.072	0.074	**0.063**	**0.060**	**0.064**	**0.063**	**0.061**	**0.063**	**0.061**	**0.062**
	0.25	MAe1	0.072	0.078	**0.070**	0.072	0.073	0.076	0.077	0.083	0.074	0.075	0.074	0.077	0.072	0.076	0.073	0.075
	0.25	MAe9	0.099	0.083	0.099	0.086	0.091	0.082	0.093	0.081	0.100	0.089	0.097	0.097	0.094	0.092	0.094	0.090
	0.25	MAe10	0.093	0.077	0.098	0.075	0.086	0.072	0.085	0.080	0.092	0.086	0.093	0.092	0.087	0.091	0.088	0.089
	0.5	NoMA	0.071	0.071	**0.069**	**0.067**	0.071	0.072	0.074	0.078	**0.063**	**0.063**	**0.066**	**0.065**	**0.064**	**0.064**	**0.065**	**0.063**
	0.5	MAe1	0.079	0.087	0.080	0.084	0.076	0.086	0.082	0.091	0.082	0.088	0.080	0.087	0.081	0.086	0.082	0.087
	0.5	MAe9	0.108	0.102	0.110	0.106	0.100	0.097	0.100	0.095	0.112	0.110	0.117	0.115	0.111	0.109	0.112	0.110
	0.5	MAe10	0.107	0.101	0.116	0.103	0.100	0.095	0.097	0.091	0.109	0.106	0.114	0.114	0.108	0.112	0.108	0.109

Table 4.10: Small Sample Within-Equation Regressor Significance Test Size Kernel and Bandwidth Calculation Comparison

	Regressor AR Coef.	VMA Process Ref.	Newey & West (1987) Manual Bandwidth								Andrews (1991) Automatic Bandwidth							
			HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)		HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)	
			α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1
No-HET	0.25	NoMA	**0.068**	0.076	**0.067**	0.073	0.072	0.078	0.074	0.084	**0.063**	**0.070**	**0.063**	**0.070**	**0.064**	0.071	**0.064**	**0.070**
	0.25	MAe1	0.082	**0.070**	0.084	**0.067**	0.080	**0.069**	0.082	0.074	0.086	**0.066**	0.085	**0.067**	0.084	**0.067**	0.084	**0.066**
	0.25	MAe9	0.102	0.083	0.104	0.082	0.092	0.082	0.090	0.081	0.100	0.089	0.099	0.092	0.098	0.090	0.098	0.089
	0.25	MAe10	0.124	0.073	0.135	0.074	0.116	**0.069**	0.105	**0.070**	0.111	0.079	0.105	0.085	0.102	0.082	0.102	0.082
	0.5	NoMA	**0.068**	0.079	**0.067**	0.075	0.074	0.081	0.076	0.085	**0.063**	**0.067**	**0.065**	**0.070**	**0.064**	**0.068**	**0.063**	**0.068**
	0.5	MAe1	0.086	0.077	0.083	0.078	0.085	0.075	0.085	0.075	0.086	0.077	0.082	0.079	0.084	0.077	0.084	0.076
	0.5	MAe9	0.109	0.107	0.115	0.110	0.103	0.100	0.100	0.099	0.116	0.109	0.119	0.110	0.113	0.109	0.111	0.108
	0.5	MAe10	0.125	0.100	0.137	0.108	0.115	0.091	0.109	0.087	0.118	0.107	0.116	0.105	0.114	0.103	0.113	0.102

Notes (both tables): Sample size of 100 observations with 2,000 replications. 'TH' refers to the Tukey-Hanning kernel, and 'QS', the Quadratic-Spectral kernel. Empirical sizes in the range 3% to 7% are shown in bold. All results refer are tests of significance of the column headed coefficient from equation (31) with respect to the true values set. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. All within-equation error processes are homoskedastic. Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

4.1.3 Summary of Empirical Size Properties

The size properties of system-HAC are not fundamentally different from those of single-equation HAC. This can be demonstrated by comparing the results obtained by Smith and Yadav (1996) in their simulations involving single-equation HAC kernels and bandwidth-choices, with our results obtained for the slope-coefficient size with a non-invertible process with no heteroskedasticity and regressor correlation of 0.5. The similarity of these results supports this claim.

Regardless of the sample size, results suggest that test-size from the system-HAC methodology is generally unaffected by either the dominant roots (processes VMAe1 and MAe951) or a complex, vector, as opposed to a simple moving average process. The same is not true for other methodologies which suffer from size distortions dependent on the properties of the error process, and in the notable case of the PCSE/Parks methodology, can be severely undersized for tests of the intercept coefficient differential if there is a VMA error process.

For both the smaller and larger samples, system-HAC test size is unaffected by the choice of kernel or bandwidth methodology. Overall, system-HAC is more often correctly sized in a sample of 300 observations compared to other methods available, for the DGPs used in our experiments.

For smaller samples, there is evidence to suggest that increasing the AR regressor correlation inflates test size above the nominal level. Further research on whether prewhitening could be used to reduce the effect of this correlation, prior to implementing system-HAC, would therefore be beneficial.

4.2 Power

Test power is evaluated by calculating the mean rejection rate over all replications for tests where the underlying DGP parameters are altered so that the difference in the parameters between equations is different from the null hypothesis (previously of the true parameter value) by ± 0.5. For intercept coefficient test power, the actual DGP coefficients are altered from the null-hypothesis value tested (against a two-tailed non-equality alternative) shown in (33) to:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 2 \end{bmatrix} \text{ therefore } \tau_1 = 1.5 \text{ (previously 1)} \qquad (37)$$

and for the slope coefficient power:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \text{ therefore } \tau_2 = 0.5 \text{ (previously 1)} \qquad (38)$$

Power is related to this proportionate distance from the true parameter value, and so for a large proportionate distance test power will be close to one for all DGPs. Likewise, for smaller proportionate distances, test power will reduce towards the empirical size of the test, and 0.5 is chosen as it offers a compromise between these two extremes.

Power is adjusted for size distortions by replacing asymptotic critical values with sample percentile counterparts. Power is only reported for the various DGPs where the corresponding test size, discussed in the previous section, is in the range 3% to 7% and so close to the nominal size of 5%, as the power of tests with size outside this range would be unachievable without further data-dependent simulation to uncover the underlying distribution of the test-statistic.

Table 4.11 shows size-adjusted power for tests of cross-equation restrictions using a sample of 300 observations. In all cases test power is influenced by the underlying DGP. For the system-HAC methodology, with no heteroskedasticity, test power is lowest when all the roots of the (V)MA process are close to non-invertibility (MAe9). Processes with one root close to non-invertibility (VMAe1, MAe951) do not have power properties which are dominated by this root, with performance more in common with processes containing the average root of those roots (MAe5). A non-invertible process (MAe10) has superior power compared to a process which is near non-invertible in all roots (MAe9). VMA processes generally produce tests with lower power than those from simple MA processes. Heteroskedasticity reduces test power for all methodologies, markedly so if the heteroskedasticity is of a complex form, HET(2), and in some cases, this test power is no larger than the empirical size.

Table 4.11: Large Sample Empirical Size-Adjusted Power of Between-Equation Tests

	Regressor AR Coef.	VMA Process Ref.	PCSE/SUR		PSCE/Parks		System-HAC (QS)	
			τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
No-HET	0.25	NoMA	0.972	1.000	1.000	1.000	0.972	1.000
	0.25	MAe1		0.999	0.997	1.000	0.925	1.000
	0.25	MAe9					0.131	0.243
	0.25	MAe10				1.000	0.314	0.659
	0.25	VMAe1				1.000	0.498	0.806
	0.25	VMAe2					0.306	0.525
	0.5	NoMA	0.856	1.000	0.984	1.000	0.853	1.000
	0.5	MAe1			0.953	1.000	0.718	1.000
	0.5	MAe5				1.000	0.215	0.648
	0.5	MAe9				0.908	0.091	0.224
	0.5	MAe10				1.000	0.193	0.574
	0.5	MAe951			0.471	1.000	0.230	0.705
	0.5	VMAe1				0.999	0.278	0.773
	0.5	VMAe2				0.983	0.154	0.522
HET(1)	0.25	VMAe1				0.970	0.291	0.531
	0.25	VMAe2				0.821	0.185	0.330
	0.5	VMAe1				0.951	0.172	0.555
	0.5	VMAe2				0.807	0.099	0.338
HET(2)	0.25	VMAe1	0.089		0.069	0.250	0.101	
	0.25	VMAe2	0.055		0.065	0.092	0.065	0.084
	0.5	VMAe1	0.070			0.109	0.075	
	0.5	VMAe2	0.050			0.056	0.045	

Notes: Sample size of 300 observations with 2,000 replications. Empirical size-adjusted power is only reported where the corresponding test size in the range 3% to 7%. All results refer to testing the equality of τ_1 or τ_2 in (31) with the originally set coefficient value. HAC results obtained using a modified Andrews (1991) 'automatic' bandwidth calculation. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. Heteroskedasticity forms relate to equation (36). Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

Comparing test power achieved by the AR(1) approximation method, PCSE/Parks, to that achieved by system-HAC, better power is generally achieved for both sets of coefficient restrictions (when correctly sized). As noted previously, however, the PCSE/Parks method suffers from inaccurate test sizing in tests of the intercept differential for many DGPs, and so power, even though sometimes low for the system-HAC methodology, is preferable. Given that the exact properties of the underlying DGP are not known by the researcher, system-HAC tests have at least some power using asymptotic critical values, regardless of the underlying DGP for nearby-alternatives versus other methodologies, which do not, as demonstrated by the blanks in Table 4.4 in terms of intercept tests for the PCSE/Parks and PCSE/SUR methods.

For the DGP which contains a weak MA processes in the errors (MAe1), with weak regressor correlation (0.25), the performance of system-HAC is near equivalent to that of the AR(1) approximation method.

Table 4.12 extends this power performance evaluation by examining the size-adjusted power properties for within-equation hypothesis tests, using a large sample (300 observations). It shows that similar test power performance is observed for these within-equation tests as was observed for between-equation tests (Table 4.11). In some cases, system-HAC power is not reported, especially when the DGP contains a strong MA process with a strong regressor correlation, as size for these particular tests were found to be outside the permissible (3-7%) range. In some of these cases, the PCSE/Parks methodology also had test size outside this range which suggests that for some DGPs all methods produce tests which are unable to correctly distinguish from nearby alternatives. Further research is required to address this issue.

Overall, system-HAC offers greater or at least comparable test power in Table 4.12 compared to other methodologies where tests are concerned with the intercept coefficient. As noted previously, all methods have reduced test power for the two-VMA processes analysed. System-HAC performance is poorer with the VMAe2 DGP versus the VMAe1 process. This deterioration in test power (for all methods) is further compounded by heteroskedasticity especially when this is of complex, HET(2) form.

Table 4.12: Large Sample Empirical Size-Adjusted Power of Within-Equation Tests

	Regressor AR Coef.	VMA Process Ref.	PCSE/SUR		PCSE/Parks		System-HAC (QS)	
			α_1	β_1	α_1	β_1	α_1	β_1
No-HET	0.25	NoMA	0.999	1.000	1.000	1.000	0.999	1.000
	0.25	MAe1		0.999	0.999	1.000	0.989	1.000
	0.25	MAe9					0.161	0.247
	0.25	MAe10				1.000	0.415	0.627
	0.25	VMAe1				1.000	0.390	0.786
	0.25	VMAe2			0.239	0.974	0.209	0.500
	0.5	NoMA	0.986	1.000	1.000	1.000	0.982	1.000
	0.5	MAe1			0.993	1.000	0.920	1.000
	0.5	MAe5				1.000	0.312	
	0.5	MAe9				0.934		
	0.5	MAe10				1.000	0.272	
	0.5	MAe951			0.470	1.000	0.345	0.662
	0.5	VMAe1				0.999		
	0.5	VMAe2			0.269	0.979		0.531
HET(1)	0.25	VMAe1				0.960	0.223	0.524
	0.25	VMAe2					0.128	0.306
	0.5	VMAe1			0.257	0.959	0.156	
	0.5	VMAe2				0.831		0.323
HET(2)	0.25	VMAe1	0.082		0.060	0.225	0.086	
	0.25	VMAe2	0.075		0.057	0.099	0.071	
	0.5	VMAe1	0.071			0.109	0.077	
	0.5	VMAe2	0.051			0.068	0.050	

Notes: Sample size of 300 observations with 2,000 replications. Empirical size-adjusted power is only reported where the corresponding test size in the range 3% to 7%. All results refer to testing the equality of α_1 or β_1 in (31) with the originally set coefficient value. HAC results obtained using a modified Andrews (1991) 'automatic' bandwidth calculation. VMA process references refer to the VMA(3) processes with parameters in Table 4.1. Heteroskedasticity forms relate to equation (36). Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities in Gauss 6.0.

Table 4.13 and

Table 4.14 present empirical test power for the various MA DGPs dependent on the bandwidth calculation method employed and the kernel used. Ignoring small differences due to the simulation methodology, there appears to be no clear difference in the test power achieved using either bandwidth methodology or from the various kernels.

Table 4.13: Large Sample Between-Equation Regressor Significance Test Size-Adjusted Power Kernel and Bandwidth Calculation Comparison

	Regressor AR Coef.	VMA Process Ref.	Newey & West (1987) Manual Bandwidth								Andrews (1991) Automatic Bandwidth							
			HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)		HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)	
			τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2
No-HET	0.25	NoMA	0.970	1.000	0.971	1.000	0.970	1.000	0.969	1.000	0.971	1.000	0.971	1.000	0.972	1.000	0.972	1.000
	0.25	MAe1	0.923	1.000	0.927	1.000	0.924	1.000	0.922	1.000	0.926	1.000	0.923	1.000	0.925	1.000	0.925	1.000
	0.25	MAe9		0.243		0.242	0.142	0.247	0.140	0.238	0.131	0.245	0.130	0.236	0.131	0.241	0.131	0.243
	0.25	MAe10						0.670	0.316	0.674	0.311	0.660	0.309	0.655	0.306	0.657	0.314	0.659
	0.5	NoMA	0.855	1.000	0.857	1.000	0.856	1.000	0.850	1.000	0.850	1.000	0.853	1.000	0.853	1.000	0.853	1.000
	0.5	MAe1	0.717	1.000	0.718	1.000	0.719	1.000	0.716	1.000	0.710	1.000	0.712	1.000	0.714	1.000	0.718	1.000
	0.5	MAe9		0.226		0.229	0.107	0.228	0.104	0.235	0.097	0.236	0.092	0.234	0.089	0.231	0.091	0.224
	0.5	MAe10							0.203	0.627				0.561	0.193	0.569	0.193	0.574

Table 4.14: Large Sample Within-Equation Regressor Significance Test Size-Adjusted Power Kernel and Bandwidth Calculation Comparison

	Regressor AR Coef.	VMA Process Ref.	Newey & West (1987) Manual Bandwidth								Andrews (1991) Automatic Bandwidth							
			HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)		HAC (Bartlett)		HAC (Parzen)		HAC (TH)		HAC (QS)	
			α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1	α_1	β_1
No-HET	0.25	NoMA	0.999	1.000	0.999	1.000	0.999	1.000	0.999	1.000	0.999	1.000	0.999	1.000	0.999	1.000	0.999	1.000
	0.25	MAe1	0.988	1.000	0.989	1.000	0.988	1.000	0.989	1.000	0.988	1.000	0.988	1.000	0.989	1.000	0.989	1.000
	0.25	MAe9				0.251	0.165	0.248	0.163	0.244	0.157	0.245	0.159	0.247	0.159	0.247	0.161	0.247
	0.25	MAe10						0.647					0.417		0.420	0.625	0.415	0.627
	0.5	NoMA	0.979	1.000	0.981	1.000	0.979	1.000	0.979	1.000	0.983	1.000	0.983	1.000	0.983	1.000	0.982	1.000
	0.5	MAe1	0.921	1.000	0.924	1.000	0.921	1.000	0.919	1.000	0.921	1.000	0.920	1.000	0.920	1.000	0.920	1.000
	0.5	MAe9							0.114									
	0.5	MAe10							0.281						0.278		0.272	

Notes (both tables): Sample size of 300 observations with 2,000 replications, with only size-adjusted power where test size in the range 3% to 7% shown. All results refer to testing equality of the column headed coefficient from equation (31) with the originally set coefficient value. 'TH' refers to the Tukey-Hanning kernel, 'QS' to the Quadratic-Spectral kernel. VMA process references refer to the VMA(3) processes with parameters defined in Table 4.1 and associated characteristic equation roots as given in Table 4.2. All within-equation error processes are homoskedastic. Regressor AR coefficient refers to the coefficient in equation (35). All results obtained by coding all methodologies and using the random-number generation facilities for the DGPs in Gauss 6.0.

Empirical size-adjusted power results for the smaller sample size of 100 observations do not offer much scope for comparison between methods, as empirical sizes are often outside the range of 3-7% for all tests with all methods except tests concerning the slope coefficient using the PCSE/Parks method.

4.2.1 Performance Evaluation Summary

System-HAC has greatest advantage over other methods when the sample size is moderately large (300 observations). Unlike other methods, system-HAC produces tests which are correctly sized for both the intercept and slope coefficients, regardless of whether these involve within or between equation restrictions on these coefficients. Although power is slightly lower in terms of the slope coefficient for the system-HAC method compared with the AR(1) method, system-HAC is able to distinguish between nearby alternatives for many more DGPs, especially when concerning cross-equation restrictions.

In small samples all methods are somewhat poor, and offer variable size properties dependent on the underlying DGP, with all methods having an inability to distinguish from nearby alternatives in some cases.

Heteroskedasticity, especially when of a complex form involving the product of the non-constant regressors is found to cause a deterioration in test power for all methods.

Bandwidth and kernel choice appears to have little effect on test power and size in large samples. This suggests that using the more simple ‘manual’ bandwidth method when the order of the process is known (although this is often not the case) has no marked performance cost.

5 An Analysis of Gender and Education Attainment Consumer Sub-Groups Inflation Expectations using System-HAC

Characteristics of consumer inflation expectations are a topical and relevant subject as Central Banks study (and perhaps attempt to manipulate) such measures of inflation perceptions to achieve monetary policy objectives.

The 'Survey of Consumer Attitudes and Behavior' conducted by the Survey Research Centre (SRC) at the University of Michigan produces aggregate US consumer inflation expectation data, constructed from approximately 500 monthly telephone interviews[22]. Respondents are asked to provide a forecast for annual inflation over the 12 months following the interview. At a monthly frequency, this data is available both for all-agents, and disaggregated into interviewee demographic sub-sets, (that is, only one demographic characteristic is constant), by gender, education, age, region of residence or household income.

Analysis of the SRC all-agent expectations data by, for example, Mankiw *et al* (2003)[23] and Baghestani (1992), has considered the question of forecast rationality and bias. This application extends that work by considering whether disaggregated expectations uncover characteristics of the response groups, which are otherwise hidden in the aggregated data. This application is limited to an analysis of the gender and education panels, building on the work Bryan and Venkatu (2001a, 2001b).

Table 5.1 summarises the categories available within each group, and the category abbreviations used in the subsequent analysis. The table also lists the base category selected when the regressor matrix used is non-block-diagonal as in (4), resulting in all other category coefficients measuring the difference from the base category.

---

[22] Approximately 60% of which are new to the survey, and 40% are reinterviews.

[23] The finding of bias in aggregate expectations by Mankiw *et al* (2003) appears incorrect when the test is repeated, with bias in the SRC survey being insignificant for all-agents. It should be noted that to achieve identical coefficient and standard-error estimates, Intercooled Stata 8.2 controls need to be overridden to allow analysis with the non-constant time horizon (quarterly data for part of the sample-period) used in the Mankiw data-set.

Table 5.1: Available Respondent Categories

Sample Split	Base Category	Remaining Categories
Gender	Male (M)	Female (F)
Education	High school degree (EHS)	Some college (ESC), College degree (ECD)

Note: all categories are pre-defined by the SRC. Average sample size information for each category is provided in the appendix.

To calculate forecast bias and rationality it is necessary to compare the forecast, $E_{i,t}\pi_{t+12}$, where $E_{i,t}$ is used to denote the expectation of group *i* formed at time *t*, with the actual realised inflation rate (12 months hence), denoted π_{t+12}. Subtracting the realised rate from the expected rate, $\pi_{t+12} - E_{i,t}\pi_{t+12}$, will produce a series of macro-economic 'centred' expectations, with over-forecasting producing a negative value, and under-forecasting producing a positive value. As such, this series can be used to quantify the forecast bias.

Inflation experience is calculated using the annual percentage change in the BLS 'Consumer Price Index Research Series Using Current Methods' (CPI-U-RS) series (see Stewart and Reed 1999 for more detailed information about this series), which will be unaffected by aggregation differences in the calculation of inflation over time and so be a temporally-consistent series representative of all urban-consumers over the entire sample period.

Following Mankiw *et al* (2003), three regression models are used to analyse potential forecast differences, with equations (groups) $i = 1 \ldots n$ being stacked into a system as discussed in Section 2:

$$\pi_t - E_{i,t-12}\pi_t = \alpha_i \tag{39}$$

$$\pi_t - E_{i,t-12}\pi_t = \alpha_i + \beta_i\left(\pi_{t-12} - E_{i,t-24}\pi_{t-12}\right) \tag{40}$$

$$\pi_t - E_{i,t-12}\pi_t = \alpha_i + \gamma_i E_{i,t-12}\pi_t \tag{41}$$

A block-diagonal system-regressor matrix, as in (3), is used since the level, and not just the difference, in forecast 'bias' of each group is relevant. Equation (39) simply regresses the centred forecast series on a constant, and α_i is equivalent to the average expectation bias for each group. Such a regression model with the block-diagonal regressor matrix set-up allows for direct tests of the significance of the bias for each group, with differences in the bias across groups being tested with post-estimation robust Wald tests. Mankiw *et al* (2003) use (40) to test for 'forecast-error persistence', while (41) is a common way of testing (weak form) 'rationality', which would hold here if $\alpha_i = \gamma_i = 0$. Again the block-diagonal regressor matrix allows for direct tests of rationality for each group, with post-estimation robust Wald tests being used to test equality of this rationality across groups.

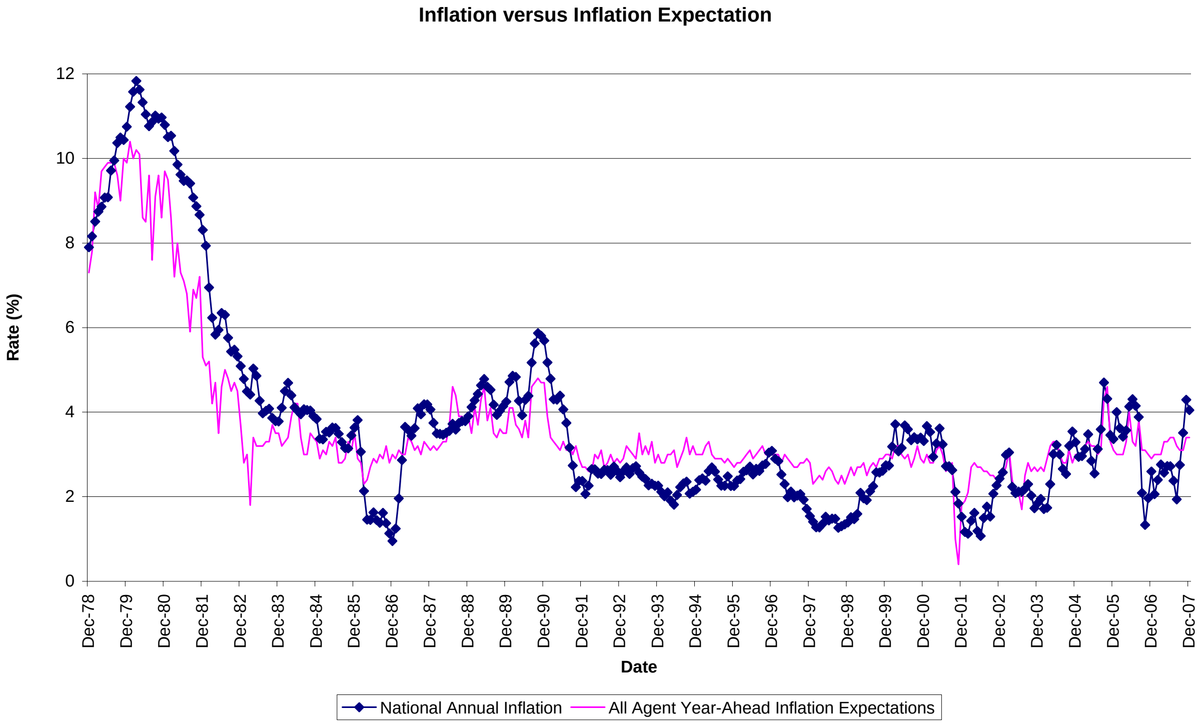
The overlapping (for eleven periods) errors in all these equations will imply a (vector)-moving average process of order 11. As such the robust-system-HAC methodology outlined in Section 3 is relevant in order to produce consistent *within* and *cross*-equation hypothesis tests. Since the order of the process is known, the 'manual' bandwidth equal to the order of the process plus one, as suggested by Newey and West (1987), is appropriate. As demonstrated by the performance evaluation results in the previous section, despite allowing the bandwidth to grow at an optimal rate with the sample size, the modified Andrews (1991) bandwidth method does not hold any clear advantage in the case of a process where the bandwidth can be approximated from the order of the process.

Note that equation (39) is a regression of the forecast error on only a constant term. Section 4 demonstrated that the benefits to using system-HAC versus other potentially robust methods were particularly strong when the researcher was interested in inference on the intercept coefficient, and so is particularly relevant to a system containing such a typical equation. Likewise, for the other equations, where inference regarding the coefficient on the intercept term coefficient is equally important, system-HAC gives a better chance of these tests being correctly sized to produce correct inference with asymptotic critical values.

Unlike Mankiw et al (2003) and Baghestani (1992) we do not consider the inflation or the mechanisms generating expectations since 1978 to have remained constant. To avoid drawing misleading conclusions from our results, it is necessary to conduct analysis on a sample with stable underlying parameters, and so a sample containing no structural breaks.

In terms of the process generating inflation it has been suggested by some commentators that Federal Reserve monetary policy has undergone a number of regime switches in response to changing economic conditions and issues of Central Bank credibility. Meulendyke (1989), for example, discusses the changing priorities placed on Federal Reserve policy objectives, and the different methods which have been employed over the years to achieve these objectives. Meulendyke (1989) and Clarida *et al* (1998) both suggest a switch in the monetary policy reaction function occurring in late 1982, following the stewardship of the Federal Reserve Chairman, Paul Volcker. The change in economic and forecasting conditions post-1982 can be seen by examining Figure 5.1 which graphs the inflation expectations against the actual rate at the time the forecast was made, that is, π_t and $E_t\pi_{t+12}$.

Figure 5.1: Inflation Expectations and Actual Inflation



It is worth noting from Figure 5.1 that consumer inflation expectations are more sophisticated than one might imagine: they both pre-empt the stabilisation and then fall in the rate of actual inflation in late-1979 and also immediately adjust forecasting processes to predict lower future inflation following the September 11th 2001 World Trade Centre terrorist attacks.

Since stable forecasting and inflation generating processes clearly do not exist across all the available data, and models are assumed to contain time-invariant parameters, the expectations data analysed here excludes the Volcker period and is truncated to start from January 1983 and is available until November 2007 (299 time-periods). For the analysis involving equations (39) to (41), as the dependent variable involves lagged expectations data, the first 12 months of data are lost, but since CPI-U-RS inflation data is available for the period up until December 2007, an extra observation is available at the end. Consequently, the analysis involving the forecast error employs 288 observations, for the period January 1984 to December 2007. Note from the previous performance evaluation analysis that by using a sample of nearly 300 observations, tests should be correctly sized using the system-HAC methodology.

5.1 Results

Table 5.2 shows the results for estimating the average bias for the gender and education equation systems, that is, estimating (39) for these groups. The penultimate column in the table presents the results of a test of coefficient equality, that is, all α_i coefficients within that system. The last column shows the test-statistic for equality of similar coefficients with zero, that is joint coefficient significance.

Table 5.2: Average Forecast Bias Results

		Coef.	Std. Error		Equality	Equality Zero
Gender	α_M	0.01	0.16		$\chi^2_{(1)}$ = 138.009***	$\chi^2_{(2)}$ =148.175 ***
	α_F	-0.39	0.15	*		
Education	α_{EHS}	-0.27	0.17		$\chi^2_{(2)}$ = 19.083***	$\chi^2_{(3)}$ 20.286***
	α_{ESC}	-0.13	0.16			
	α_{ECD}	-0.17	0.15			

Note: * denotes significance at the 5% level, ** at the 1% level, *** at the 0.1% level. A blank in this column indicates significance only at levels above 5%. Results refer to estimation of the system of equations with typical equation (39) set-up as per (3), that is, with a diagonal regressor matrix. Group-codes are as given in Table 5.1. Robust standard-errors calculated using the system-HAC methodology, calculating the system covariance matrix using (29), with a 'manual' bandwidth of 12 and a Quadratic-Spectral kernel. Post-estimation test employ this robust system covariance matrix with a robust Wald test.

Results suggest that the bias in male forecasting is small, and insignificantly different from zero. Female forecasts, however, on average over the sample period, are biased upward by 0.39%: suggesting, perhaps, that females are more pessimistic. This leads to rejection of the null-hypothesis that coefficients are equal and a rejection of the null-hypothesis of joint insignificance.

For the education system, although individual coefficients seem relatively large, no education grouping individually has statistically significant bias. The large difference between the overpredication by the high school education (EHS) and some college educated (ESC) is the probable cause of the rejection of the null-hypothesis of joint-coefficient equality and the test of joint insignificance.

These results in Table 5.2 contrast with the average all-agent forecast for the Mankiw *et al* (2003) period which suggest no-forecast bias[24].

Table 5.3 presents the results for estimating a system for forecast persistence, with typical equation given by (40).

Focusing on the coefficient on the realised forecast error, β_i, for neither the gender nor the education group is this coefficient significant either individually or jointly. Mirroring the results of the previous regression model, there is significant individual bias for the female group (independent of the scale of the realised forecast error), which jointly results in these coefficients for the gender system being significantly different from each other and from zero. For the education panel, intercept coefficients are jointly different and significant, but as with the previous analysis, are individually insignificant. Jointly, all coefficients, for both systems, are significant, as shown by the test-statistic in the last column of the table.

---

[24] Repeating the analysis on page 17 of the Mankiw *et al* (2003) paper produces identical coefficients and standard-errors with significance being correctly identified in all but this regression model, where no-stars (used to signify significant bias) should have been included.

Table 5.3: Forecast Bias Persistence Results

		Coef.	Std. Error		Equality	Equality Zero	Significance All
Gender	α_M	-0.04	0.165		$\chi^2_{(1)}$ 52.946***	$\chi^2_{(2)}$ 53.070***	$\chi^2_{(4)}$ 163.795***
	α_F	-0.49	0.175	**			
	β_M	-0.10	0.139		$\chi^2_{(1)}$ 0.306	$\chi^2_{(2)}$ 1.549	
	β_F	-0.13	0.125				
Education	α_{EHS}	-0.35	0.181		$\chi^2_{(2)}$ 21.113***	$\chi^2_{(3)}$ 21.756***	$\chi^2_{(6)}$ 70.124***
	α_{ESC}	-0.19	0.166				
	α_{ECD}	-0.22	0.165				
	β_{EHS}	0.09	0.139		$\chi^2_{(2)}$ 0.448	$\chi^2_{(3)}$ 1.500	
	β_{ESC}	-0.08	0.136				
	β_{ECD}	-0.14	0.125				

Note: * denotes significance at the 5% level, ** at the 1% level, *** at the 0.1% level. A blank in this column indicates significance only at levels above 5%. Results refer to estimation of the system of equations with typical equation (40) set-up as per (3), that is, with a diagonal regressor matrix. Group-codes are as given in Table 5.1. Robust standard-errors calculated using the system-HAC methodology, calculating the system covariance matrix using (29), with a 'manual' bandwidth of 12 and a Quadratic-Spectral kernel. Post-estimation test employ this robust system covariance matrix with a robust Wald test.

Table 5.4: Forecast Rationality Results

		Coef.	Std. Error		Equality	Equality Zero	Significance All
Gender	α_M	1.27%	0.652%		$\chi^2_{(1)}$ 0.218	$\chi^2_{(2)}$ 1.412	$\chi^2_{(4)}$ 909.766***
	α_F	1.40%	0.687%	*			
	γ_M	-0.4377	0.24357		$\chi^2_{(1)}$ 4.212	$\chi^2_{(2)}$ 7.920*	
	γ_F	-0.5453	0.22082	*			
Education	α_{EHS}	1.79%	0.725%	*	$\chi^2_{(2)}$ 1.48	$\chi^2_{(3)}$ 8.713*	$\chi^2_{(6)}$ 154.170***
	α_{ESC}	1.62%	0.570%	**			
	α_{ECD}	1.22%	0.515%	*			
	γ_{EHS}	-0.6511	0.23740	**	$\chi^2_{(2)}$ 1.686	$\chi^2_{(3)}$ 8.507*	
	γ_{ESC}	-0.5804	0.20020	**			
	γ_{ECD}	-0.4530	0.18128	*			

Note: * denotes significance at the 5% level, ** at the 1% level, *** at the 0.1% level. A blank in this column indicates significance only at levels above 5%. Results refer to estimation of the system of equations with typical equation (41) set-up as per (3), that is, with a diagonal regressor matrix. Group-codes are as given in Table 5.1. Robust standard-errors calculated using the system-HAC methodology, calculating the system covariance matrix using (29), with a 'manual' bandwidth of 12 and a Quadratic-Spectral kernel. Post-estimation test employ this robust system covariance matrix with a robust Wald test.

Conversely, Mankiw *et al* (2003), for the aggregate all-agent series, find evidence of forecast persistence, with the realised forecast error coefficient being significant. The constant term in their regression, however, is not significant.

The results of estimating (41), that is, investigating forecast rationality for the demographic groups, are shown in Table 5.4.

Again, significance in the gender panel is focussed on the female sub-sample. Interestingly, all coefficients in the education panel are individually and jointly significant (though the hypothesis that they are identical cannot be rejected).

Table 5.5 continues this analysis by testing the null-hypothesis that $\alpha_i = \gamma_i = 0$ for each equation in the system. Rejecting this null-hypothesis is equivalent to rejecting rationality, for that group. Unsurprisingly, the test rejects at the 5% level for females, inferring that females are not rational forecasters for CPI-U-RS inflation over the sample period. In contrast to the previous, more general evidence, rationality is also rejected for those only educated to high school standard, and this rejection is significant at the 1% level.

Table 5.5: Tests of Rationality

	M	F
$\chi^2_{(2)}$ test statistic	3.972	8.585*

	EHS	ESC	ECD
$\chi^2_{(2)}$ test statistic	14.700**	5.833	2.918

Note: * denotes significance at the 5% level, ** at the 1% level, *** at the 0.1% level. A blank in this column indicates significance only at levels above 5%. Results refer to testing the null-hypothesis $\alpha_i = \gamma_i = 0$ following estimation of the system of equations with typical equation (41) set-up as per (3), that is, with a diagonal regressor matrix. Group-codes are as given in Table 5.1.

For the aggregate all-agent series, Mankiw *et al* (2003) reject the null-hypothesis of forecast rationality. The present analysis, however, shows that the likely source of this aggregated irrationality comes from females or those who are only educated to high-school standard.

To further understand these forecasting differences it is useful to compare and test the significance of the differences in non-centred forecasts, that is, to compare average expectations over the sample period. This can be achieved by constructing a system with typical equation:

$$E_{i,t-12}\pi_t = \alpha_i \tag{42}$$

As discussed previously, the truncated expectations data runs from January 1983 to November 2007, resulting in 299 observations available for analysis.

A non-block-diagonal system-regressor matrix, as in (31) is more useful in this situation, as it will allow for direct test of the coefficient differences and the level of almost certainly non-zero average forecasts is of little interest.

The errors of such a model will not follow a known correlation pattern, as was previously the case when analysis concerned the centred 'forecast-error' series. Accordingly, results presented in Table 5.6 use two methodologies, the system-HAC methodology with the 'automatic' data-dependent modified Andrews (1991) bandwidth, and the AR(1) extension to the PCSE methodology. It is hoped that using two-estimation methodologies will make the results more robust to the unknown correlation structure.

Table 5.6: Average Forecast Differences

		Coef.	System-HAC Std. Error		Bandwidth	PCSE AR(1) Std Err	
Gender	α_M	2.89	0.09	***	15	0.08	***
	$\alpha_F - \alpha_M$	0.40	0.04	***		0.09	***
Education	α_{EHS}	3.18	0.10	***	19	0.06	***
	$\alpha_{ESC} - \alpha_{EHS}$	-0.15	0.05	**		0.07	*
	$\alpha_{ECD} - \alpha_{EHS}$	-0.12	0.11			0.12	

Note: * denotes significance at the 5% level, ** at the 1% level, *** at the 0.1% level. A blank in this column indicates significance only at levels above 5%. Results refer to estimation of the system of equations with typical equation (42) set-up as per (31), that is, with a diagonal regressor matrix. Group-codes are as given in Table 5.1. System-HAC robust standard-errors are calculated using the system-covariance matrix (29), with a 'automatic' bandwidth and a Quadratic-Spectral kernel. PCSE AR(1) standard errors are calculated using Intercooled Stata8.2 xtpcse command.

It is unsurprising that females' forecasts are statistically significantly different at the 0.1% level (both methodologies) from those of males, being on average 0.4% higher[25]. Likewise, compared to those educated only to high school standard, those with some college education forecast lower inflation by approximately 0.15%, which is statistically significant at the 1% level using the system-HAC methodology, and at 5% using the PCSE AR(1) methodology.

The forecasting difference for females clearly translates into the tendency to over-predict inflation, as demonstrated by the results presented previously. Bryan and Venkatu (2001a), for a different time-sample, also observe (but do not test the significance) that females have higher average expectations than males. This finding might be correlated to the exposure of females, compared to males, with a different set of prices, for example food prices, from which they gauge inflation.

Results in Table 5.5 suggest that those only high-school educated may be irrational forecasters. As such the finding in Table 5.6 that those educated with some college tutoring may forecast lower compared with those whom are only high-school educated, might be a rational response for this group. Bryan and Venkatu (2001a) also find that those educated only to high-school standard forecast the highest inflation compared to other education groupings while similar analysis for UK inflation perceptions reveals a similar trend (see Lombardelli and Saleheen 2003), despite the obvious geographic difference in the commodities being used to measure inflation. It remains to be understood why this forecasting difference for those with a college degree is insignificant and yet there is no previous evidence of agents in this group being irrational or biased forecasters.

It should be noted that although the CPI-U-RS may be an accurate gauge of inflation experience for the average consumer, it might poorly represent the inflation experience of females and/or those who are only educated to high-school standard. Forecast bias or irrationality in relation to this measure of inflation may therefore be 'rational' for these sub-groups of consumers.

---

[25] These results are different to the coefficient differences which could be calculated by hand from Table 5.2 since the sample period is slightly different in the analysis presented in Table 5.6.

6 Conclusions and Summary

Estimating systems-of-equations for time-series survey data is problematic because data dependencies can exist within and between equations. Current methods to deal with this type of data include the SUR, Parks and PCSE methodologies. However, these estimation methodologies are somewhat restrictive in the structure of system error correlation which can be accommodated. Specifically, there is no non-parametric method currently available to accommodate a vector moving average (VMA) error process in the system errors.

Extending the single-equation heteroskedasticity and autocorrelation (HAC) robust work of Newey and West (1987) this paper has shown how consistent estimates of model parameters can be produced, robust to autocorrelation and heteroskedasticity, but in particular, robust to VMA error process combined with heteroskedasticity. This system-HAC methodology, in constructing a robust proxy of the system error covariance matrix, also permits robust post-estimation tests of parameter restrictions.

Detailed simulation evidence for a fixed-order VMA process suggests that the performance of system-HAC in terms of size and size-adjusted power is good, compared with other potentially system-robust methodologies, in large samples (300 observations). Results show that for the range of data-generation processes and forms of heteroskedasticity, system-HAC in large samples is generally correctly sized, and unlike other methodologies, is generally so for tests involving the intercept coefficient. All methodologies suffer from power dependence on the underlying DGP, but unlike other methodologies, system-HAC always has some power to distinguish from nearby-alternatives, regardless of the DGP, for tests of cross-equation restrictions. Performance of all methodologies with small-samples (100 observations) tends to be poor, with all methods being unable to accurately test hypothesises against nearby-alternatives for some DGPs. Finite sample refinements to this methodology should be seen as an area for investigation in future research.

Simulations have also considered the effect of the kernel and bandwidth calculation methodology. Results suggest that both these choices have little influence on the size or power properties of system-HAC.

Further investigation of whether a block-bootstrap type approach can be used to recover the finite-sample distribution of finite-sample test statistics, and so improve finite sample performance, remains to be seen. Performance might also be improved by an appropriate system-modification to prewhitening, a topic also worthy of future research.

In constructing a system-of-equations to compare average bias in demographic sub-groups of year-ahead consumer inflation expectations data, system-errors are known to follow a (vector) moving average process. Applying system-HAC for the purposes of this analysis, expectations for females consumers are found to be biased in relation to average annual inflation, with over-forecasting of 0.39%. Despite the fact that in neither the gender nor education sub-group analysis are forecast-errors found to be persistent, using the standard test of rationality, both sub-groups are found to be forecasting irrationally.

Underlying average forecasts are found to be adrift for females from those of males by 0.4%. Likewise those with only some college education are found to be different from those only high-school educated by 0.15%. Why those educated to college-degree standard are not adrift, and yet, are not biased or irrational, remains to be understood.

The application of the system-HAC has demonstrated that demographics are an important characteristic to inflation-forecasting processes. Furthermore, it has shown that generally used aggregated data can mask interesting forecasting differences contained within sub-groups of consumers: aggregate forecasting patterns (as investigated by Mankiw et al 2003, and others) are not necessarily the same as those for sub-groups of consumers. It has also confirmed econometrically the note by Bryan and Venkatu (2001b) that men and women have “curiously different inflation perspectives”.

References

Andrews, D. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59(3 – May), 817-858.

Andrews D. (2002) The Block-Block Bootstrap: Improved Asymptotic Refinements, *Cowles Foundation Discussion Paper*, 1370, May.

Andrews, D.; Monahan, J. (1992) An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica*, 60(4 – July), 953-966.

Baghestani, H. (1992) Survey Evidence on the Muthian Rationality of the Inflation Forecasts of US Consumers, *Oxford Bulletin of Economics and Statistics*, 54(2), 173-186.

Beck, N.; Katz, J. N. (1995) What to do (and not do) with Time-Series Cross-Section Data, *The American Political Science Review*, 89(3 – September), 634-647.

Bryan, M.; Venkatu, G. (2001a) The Demographics of Inflation Opinion Surveys, *Federal Bank of Cleveland Economic Commentary*, October.

Bryan, M.; Venkatu, G. (2001b) The Curiously Different Inflation Perspectives of Men and Women, *Federal Bank of Cleveland Economic Commentary*, November.

Clarida, R.; Gali, J.l Gertler, M. (1998) Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *NBER Working Paper*, 6442

Davidson, J. (2001) *Econometric Theory*, Blackwell: Massachusetts

Davidson, R.; MacKinnon, J. (1985) Heteroskedasticity-Robust Tests in Regression Directions, *Annales de L'Insee*, 59/60, 183-217.

den Haan, W.; Levin, A. (1997) A Practitioners Guide to Robust Covariance Matrix Estimation, *mimeo.*

den Haan, W.; Levin, A. (1998) Vector Autoregressive Covariance Matrix Estimation, *mimeo*.

den Haan, W.; Levin, A. (2000) Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order, *NBER Working Paper*, 255, June.

Galbraith, J.; Zinde-Walsh, V. (1994) A Simple Noniterative Estimator for Moving Average Models, *Biometrika*, 81(1-March), 143-155.

Galbraith, J.; Ullah, A.; Zinde-Walsh, V. (2002) Estimation of the Vector Moving Average Model by Vector Autoregression, *Econometric Reviews*, 21(2), 205-219.

Godfrey, L.; Orme, C. (2003) Controlling the Finite Sample Significance Levels of Heteroskedasticity-Robust Tests of Several Linear Restrictions on Regression Coefficients, *Economic Letters*, 82, 281-287.

Hansen, L.; Hodrick, R. (1980) Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy*, 88(5 - October), 829-853

Hansen, B. (1992) Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes, *Econometrica*, 60(4 –July), 967-972.

Lin, C.; Sakata, S. (2005) Consistent Estimation of Long-Run Covariance Matrices with Truncated Flat Kernel, *mimeo*.

Lombardelli, C.; Saleheen, J. (2003) Public Expectations of UK Inflation, *Bank of England Quarterly Bulletin*, Autumn, 281-289.

MacKinnon, J. (2002) Bootstrap Inference in Econometrics, *Canadian Journal of Economics*, 35(4), 615-645.

Mankiw, N. G.; Reis, R.; Wolfers, J. (2003) Disagreement about Inflation Expectations, *NBER Working Paper*, 9796, June.

Meulendyke, A. (1989) The Federal Reserve and U.S. Monetary Policy: A Short History, in U.S. Monetary Policy and Financial Markets, Chapter 2.

Newey, W.; West, K. (1987) A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelations Consistent Covariance Matrix, *Econometrica*, 55(3 – May), 703-708.

Newey, W.; West, K. (1994) Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies*, 61, 631-653.

Nicholls, D.; Pagan, A.; Terrell, R. (1975) The Estimation and Use of Models with Moving Average Disturbance Terms: A Survey, *International Economic Review*, 16(1 – February), 113-134.

Parks, R. W. (1967) Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated, *Journal of the American Statistical Association*, 62(318 – June), 500-509.

Smith, J.; Yadav, S. (1996) A Comparison of Alternative Covariance Matrices for Models with Over-lapping Observations, *Journal of International Money and Finance*, 15(5), 813-823.

Stewart, K.; Reed, S. (1999) CPI research series using current methods, 1978-1998, *Monthly Labor Review*, June, 29-38.

Sul, D.; Phillips, P.; Choi, C. (2005) Prewhitening Bias in HAC Estimation, Oxford Bulletin of Economics and Statistics, 67(4), 517-546

White, H. (1980) A Heterokedasticity-Consistent Covariance Matrix Estimator and Direct Test for Heteroskedasticity, *Econometrica*, 48(4 – May), 817-838.

Wooldridge, J. (2002a) *Econometric Analysis of Cross Section and Panel Data, MIT: Cambridge Mass.*

Zellner, A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, *Journal of the American Statistical Association*, 57(298 - June), 348-368.