A Hypergeometric Test for Omitted Nonlinearity*

Steve Lawford†
Department of Economics and Finance, Brunel University, UK

April 24, 2003

Abstract

A frequently used test for unspecified nonlinear omissions is the parametric RESET, which is based upon a finite polynomial. We follow Abadir (1999), who suggests that the generalized hypergeometric function may provide a more flexible proxy for the omission; and propose a new approach, semi-nonparametric in spirit, that is based upon estimation of the hypergeometric parameters, and which does not require large datasets. While minimal ex ante assumptions are made about the functional form, this is fully revealed following implementation. Using Monte Carlo simulation, we examine null distributions, and then show that the small-sample power of our test can be a considerable improvement over that of the RESET, when the correct class of functional forms of the omission is known. We investigate a variety of theoretical and numerical issues (including rapid and stable numerical optimization) that arise in development of a workable procedure, and offer practical solutions that should be especially useful whenever hypergeometrics are applied to problems of modelling nonlinearity.

Keywords: Hypergeometric functions; Monte Carlo simulation; Numerical optimization; Omitted variables; RESET test
JEL classification: C12; C13; C20; C63; C65

---

*I am indebted to Karim Abadir, for his tireless advice and encouragement. For useful discussions during the development of this work, I thank Giovanni Forchini, Les Godfrey, Mohammad Kashani, Patrick Marsh and Michael Rockinger, and seminar participants at Université Libre de Bruxelles and Universidade da Madeira. I am very grateful to Marc Hallin and ECARES for their hospitality during 2000–2002. Financial support was provided by a University of York Teaching Fellowship and a Marie Curie Individual Fellowship under the European Commission's Improving Human Potential programme, and is gratefully acknowledged. The author is solely responsible for information communicated. This paper was typed in Scientific Word 3.0 and numerical results were derived using GAUSS.

†Department of Economics and Finance, Brunel University, Uxbridge, UB8 3PH, UK; steve_lawford@yahoo.co.uk; fax +44 1895 203384.

1 Introduction

A new and potentially rewarding application of the generalized hypergeometric function, or ${}_pF_q$, that has barely been considered in any field, is its use for modelling nonlinearity, via estimation of the parameters of the hypergeometric. This methodology is introduced in a seminal paper by Abadir (1999), who suggests methods by which the ${}_pF_q$ may be applied to problems of econometric modelling and testing, by exploiting the flexibility of its general nonlinear character. Some of the ideas therein are extended by Lawford (2001) and Abadir and Rockinger (2003), and are illustrated in simulations and an application respectively; we discuss their findings in further detail below. The main advantages of the general methodology are as follows: while minimal ex ante assumptions are made about the functional form, this is fully revealed following implementation, in a parsimonious manner (hypergeometric functions cover a wide variety of classes of functional form, e.g. exponential, binomial, depending upon the choice of parameter values, while the particular values assumed by these parameters will alter the shape of the function); and the estimation procedure does not require large datasets, so avoiding certain difficulties associated with fully nonparametric methods.

The purposes of this paper are two-fold. Firstly, we motivate our development of a workable procedure by focusing on the specific problem of testing for omitted nonlinearity in a static regression model, and present a new, and statistically powerful, destructive misspecification test. Secondly, we investigate a variety of theoretical and numerical issues, (including rapid hypergeometric generation and stable numerical optimization), in the testing framework. We highlight some of the reasons for the previous success of the ${}_pF_q$ in theoretical settings, (e.g. great flexibility and generality, with important special cases; the feasibility of efficient numerical generation; and access to a body of mathematical relationships between ${}_pF_q$'s), which are equally valuable when designing a constructive modelling (or testing) procedure. The practical advice that we offer should be useful when hypergeometrics are applied to more general problems of modelling nonlinearity.

1.1 A motivating problem

When testing for unspecified nonlinear omissions, Ramsey and Schmidt's (1976) Regression Specification Error Test (RESET or "RESET test") is often implemented. A parametric test, it fits a polynomial function to the omitted variable; hence, the power of the RESET depends upon the correlation of the polynomial with the actual misspecification, and upon the number of terms included within that polynomial. It has experienced widespread use in economics since the early applied papers of Loeb (1976) [quarterly investment models] and Ramsey and Alexander (1984) [business cycle analysis], and is implemented within many modern econometric software packages, including Microfit, PC-Give, SHAZAM, and TSP.

It is important to check that a regression model is correctly specified and that no relevant regressors have been omitted. These violations are generally assumed to have a detrimental effect upon the model performance: omission of relevant variables will usually lead to inconsistency of parameter estimators, thus rendering standard inference invalid. As Ramsey (1983, p. 244) notes, "one needs a general, moderately robust, information parsimonious procedure for a specification error test for omitted variables or incorrect functional form"; and this can be a very useful tool for the detection of inadequate models (see Godfrey, 1988, and Davidson, 2000, for discussion of "destructive" misspecification testing). The usefulness of general tests for misspecification motivates the search for a more flexible proxy for the omission than the finite polynomials used by RESET, which is not too costly in terms of degrees of freedom, and which is applicable to small datasets.

1.2 The plan

The structure of this paper is as follows: In Section 2, we briefly review the reasoning underlying the RESET test, which we use later in the paper, as a benchmark against which our new test will be compared. Since hypergeometric functions have not yet been adopted as standard tools in economic research we discuss, in Section 3, their successful use in a variety of fields, and then outline the relevant theory of special functions and detail efficient computer generation algorithms: comprehensive GAUSS code was created, which is a useful contribution in itself and a vital component of a workable procedure. We do not attempt an exhaustive survey but detail only such theory as will be useful later in this paper. Hypergeometric functions are then combined with a numerical optimization technique and used in a new and promising manner in Section 4, in order to derive a new test for additive omitted (or "neglected") nonlinearity, with a discussion of the computational difficulties involved. Unlike the RESET, we do not restrict the functional form of the omission to be a particular polynomial: instead, we allow it to be a member of a class of hypergeometric functions which can include many important special cases.

In Section 5, we present the main results from an extensive Monte Carlo simulation study, where the small-sample performance of the hypergeometric test is appraised and compared with two variants of the RESET test. We examine null distributions, and then show that the small-sample power of our test can be a considerable improvement over that of the RESET, when the correct class of functional forms of the omission is known. Since our test is based upon numerical methods, we consider some issues of robustness. Although we have chosen misspecification testing to illustrate the methodology, we offer practical solutions that should be useful whenever hypergeometrics are applied to problems of nonlinear modelling.

In Section 6, we briefly discuss extensions of this work in two main directions. Firstly, to a parsimonious generalized hypergeometric test for omitted nonlinearity, given that the class of functional forms of the omission is unknown, possibly by exploiting the confluences (mathematical links) between ${}_pF_q$'s of different order, in a general-to-specific manner; see Abadir (1999). Secondly, an application of related techniques to a constructive modelling approach, generalizing the Box-Cox transformation (Box and Cox, 1964), whereupon the estimated hypergeometric parameters may be interpretable to some extent. The desired theoretical foundation for all of this work (e.g. proofs of consistency) is also mentioned. We include details of the speed and accuracy of our vector hypergeometric generation code in an Appendix.

The new test is semi-nonparametric (or pseudo-nonparametric) in the sense that minimal assumptions are made about the underlying parameterization of the omission.[1] It attempts to deduce relationships using the data alone. However, this definition is not strict, since the functional form of the omission is fully revealed following implementation, potentially indicating the type and characteristics of the nonlinearity. This paper contributes to an "illustrated general theory for estimation without prior knowledge of functional forms, by means of the generalized hypergeometric series [that] is currently being developed by Abadir, Lawford and Rockinger." (Abadir, 1999, p. 295). While nonparametric approaches to explain data do exist, they do not generally reveal any structural details of the relationship. An alternative approach of modelling, estimation and testing – and particularly one that will parsimoniously reveal the structural details of nonlinear re-

[1] "The prefix semi means half and the term seminonparametric (SNP) . . . is intended to convey the notion that SNP procedures are halfway between parametric and nonparametric inference procedures." (Gallant and Tauchen, 1989, p. 1093). The most commonly applied SNP method is based upon Hermite polynomials and consists of modelling conditional density functions as nonlinear series expansions: the leading term is chosen to be a parametric model, while higher-order terms accommodate deviations from the leading term (and capture complicated structure in the data). The approach solves certain dimensionality problems commonly associated with fully nonparametric methods. For applications, see Gallant, Rossi and Tauchen (1992) and references in Podivinsky (1996).

lationships – is possible; and this may be compared with other successful techniques, such as artificial neural networks (e.g. Kuan and White, 1994), and those discussed in Granger and Teräsvirta (1993).

When existing methods are considered, our approach is perhaps closest to fitting a system of orthogonal Hermite polynomials; e.g. see Madan and Milne (1994) and Abken, Madan and Ramamurtie (1996) for work on approximating risk-neutral densities using orthogonal Hermite polynomials. In this paper, we estimate the parameters of hypergeometric functions rather than systems of polynomials. In work that is directly related to ours, Abadir and Rockinger (2003) estimate the parameters of hypergeometrics (${}_1F_1$'s) in a modelling framework. They use a mixture of Kummer functions, with parameter restrictions, to estimate density-related functionals where no prior knowledge of the underlying functional form is available, and when the variate may not be directly observable. Their method is successfully applied to problems in option pricing (French Franc/DeutschMark European exchange rate options), given small datasets (e.g. 10-20 observations). We face a different set of problems in this paper since we seek a testing procedure that is, to some extent, automated; and that is efficient enough for simulation studies to be a practical option. As a result, it becomes more difficult to exploit certain properties of the hypergeometrics, e.g. argument transformations.

Lawford (2001, chs. 2 and 3) develops some of the theoretical and computational aspects of modelling nonlinear relations using hypergeometric functions, where the correct class of functional forms is known; and is the first instance of use of a variety of ${}_pF_q$'s in a general applied manner, and to testing and simulation work rather than modelling. In this paper, we estimate the parameters of a variety of hypergeometrics: ${}_0F_0$, ${}_1F_0$, ${}_0F_1$, ${}_0F_2$, ${}_1F_1$, ${}_1F_2$, ${}_2F_1$, some of which require particular care in evaluation and estimation (e.g. ${}_1F_1$, ${}_2F_1$). Although we find that it is currently too time-consuming to perform simulations based upon the ${}_2F_1$, we illustrate the potential of the efficient generation procedure in the Appendix, with an application of the ${}_2F_1$ to constructive modelling.

Other related research is by Gordy (1998a), who introduces the compound confluent hypergeometric (CCH) distribution, with an empirical application involving modelling of the distribution of measures of household liquid assets across households. The CCH involves a confluent hypergeometric function of 2 variables (e.g. see Gradshteyn and Ryzhik, 1980), and generalizes a variety of other distributions that have commonly been used in the statistical modelling of bounded random variables, e.g. beta; and the confluent hypergeometric distribution (CH) (Gordy, 1998b); see also McDonald (1984). This work is motivated by similar considerations to our paper, and the aim is to allow for a more flexible description of data while imposing more structure than that offered by a nonparametric estimator. The CCH can be rapidly calculated over most of the parameter space, and can take the U-shaped and single-peaked forms of the beta pdf, and also a variety of multi-modal and long-tailed forms. Estimation of the parameters of the CCH is performed by maximum likelihood, and the CCH is shown to offer greater flexibility in fitting data than previous methods, visible differences in fit, and additional precision that is statistically significant.

A similar approach to that of Gordy is employed by Al-Saqabi, Kalla and Tuan (2003), who develop a generalized gamma-type pdf; and in a different context by Kumar (2002), who introduces a new class of discrete distributions, termed extended generalized hypergeometric probability distributions (EGHPD), as a generalization of generalized hypergeometric probability distributions – although some properties are derived analytically, (e.g. moments, moment generating function, and hazard rate), there is no discussion of parameter estimation. The ${}_pF_q$ also plays a central role in the work of Gottschling, Haefke and White (1999), who derive a new family of analytically tractable and flexible (log-)hypernormal pdf's, based upon the logarithm of the inverse Box-Cox transform; and use techniques of artificial neural networks to yield arbitrarily accurate approximations to classes of functions whose antiderivatives have closed-form expressions for their integrals; and in Giacomini, Gottschling, Haefke and White (2002).

Notation: We generally follow the suggestions on notation in Abadir and Magnus (2002); although we differ in our use of $\overline{a}$ to indicate that a parameter is fixed during a numerical optimization, and $a^\star$ to represent an optimized parameter. Throughout this paper, we represent scalar, vector and matrix quantities as a, $\boldsymbol{a}$ and $\boldsymbol{A}$ respectively; these have representative elements $\boldsymbol{a} = \{a_j\}$ and $\boldsymbol{A} = \{a_{ij}\}$. Special examples include the $k \times 1$ zero vector $\mathbf{0}_k$ and the $k \times k$ identity matrix $\boldsymbol{I}_k$. The sets of reals and integers (including zero) are denoted by $\mathbb{R}$ and $\mathbb{Z}$ respectively, where subscript $+, -$ indicate subsets of positive and negative numbers; e.g. $\mathbb{Z}_+ = \{1, 2, \dots\} \equiv \mathbb{N}$, the set of natural numbers. We introduce new notation as needed.

2 The RESET test

The RESET test was proposed by Ramsey (1969), following discussion of graphical methods of residual analysis by Anscombe (1961) and Anscombe and Tukey (1963). It is a general test for misspecification, which is designed to detect both omitted variables and incorrect functional form by testing for a non-zero conditional mean of the disturbances, against the alternative of specification error. The RESET is designed to handle situations in which one has already incorporated all available a priori information, i.e. all variables thought to be most relevant.[2] Consider the null model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}\left(\boldsymbol{0}_N, \sigma^2 \boldsymbol{I}_N\right), \quad 0 < \sigma^2 < \infty, \tag{1}$$

such that $\boldsymbol{y}$ is an $N \times 1$ vector of observations on the dependent variable; $\boldsymbol{X}$ is an $N \times K$ non-stochastic matrix, with full column rank, of known observations on K explanatory variables: it will be assumed that the first column of $\boldsymbol{X}$ is an $N \times 1$ vector of ones; $\boldsymbol{\beta}$ is a $K \times 1$ vector of unknown (and unobservable) real coefficients; $\boldsymbol{u}$ is an $N \times 1$ vector of disturbance terms, independently and identically normally distributed.

The fundamental assumption of RESET is that some unknown analytic function of $\boldsymbol{X}\boldsymbol{\beta}$ provides a good approximation to the omitted factor. A polynomial approximation to this function is used and $\boldsymbol{\beta}$ replaced by the ordinary least squares (OLS) estimate from (1). The procedure may be formulated as a variable addition test, by testing the significance of some matrix of regressors $\boldsymbol{Z}$ in the augmented regression

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{v}, \tag{2}$$

where $\boldsymbol{Z}$ is an $N \times s$ matrix of explanatory variables and $\boldsymbol{\gamma}$ is an $s \times 1$ vector of parameters; $\boldsymbol{v}$ is an $N \times 1$ vector of disturbance terms. The greater the correlation between $\boldsymbol{Z}$ and the nonlinear part of the true conditional mean of $\boldsymbol{y}$, the greater (in general) will be the power. Although any test variables which are correlated with the omitted variables will lead to a test with some power against the alternative, Ramsey (1969) suggested that $\boldsymbol{Z}$ comprise powers of the fitted $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, so that $(\boldsymbol{Z})_{ij} = \{\widehat{y}_i^{j+1}\}$, and this has become the standard choice. Ramsey's original presentation uses Theil's

[2] If a particular model is inadequate due to specification error, there should be some indication in the residuals (the distribution will differ from that postulated under the null). Unless the regressors and the omitted variable(s) are orthogonal, the residuals will have a non-zero expectation. It should then be possible to detect some residual pattern either visually or, more rigorously, by means of an appropriate test statistic.

(1965, 1968) BLUS residuals. The simpler approach that is now widely used and that is implemented in this paper uses OLS residuals and derives from his later work (e.g. Ramsey and Schmidt, 1976); this test is equivalent to the original BLUS formulation.

The null hypothesis $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ (no omitted nonlinearity) is tested against the two-sided alternative $H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$ using the standard F test

$$\mathrm{F}_R = \left(\frac{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{\widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}}} - 1\right)\left(\frac{N-K-s}{s}\right) \underset{H_0}{\sim} \mathrm{F}\left(s, N-K-s\right), \tag{3}$$

where $\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{v}}'\widehat{\boldsymbol{v}}$ are the sums of squared residuals from (1) and (2) respectively, and the distribution of F_R is known exactly under the null.[3] The statistic F_R is known to have an approximate doubly non-central F distribution under the alternative hypothesis. We reject the null hypothesis at the $100\alpha\%$ level of significance as $\mathrm{F}_R > \mathrm{F}^{\alpha}(s, N-K-s)$. We consider the conventional 5% level of significance in our simulation work.[4]

The empirical implementation of the RESET test involves choosing the highest power, $(J+1)$, of $\widehat{y}_i$ to be used in the auxiliary regression. Since the value of J is irrelevant to the exactness of the RESET, its importance lies in its impact on power. As we include more variables, degrees of freedom are lost, which could considerably reduce the power of the test. The first power of $\widehat{y}_i$ is omitted since the predicted value would be perfectly correlated with the K regressors of the original model (i.e. $\widehat{\boldsymbol{y}}$ collinear with $\boldsymbol{X}$). The properties of the RESET were assessed by Ramsey and Gilbert (1972), who provide favourable evidence on the small-sample performance of the RESET in simulations, with $J = 3$; and by Thursby and Schmidt (1977). We will also consider $\boldsymbol{Z} = (\boldsymbol{z})_i = \{\widehat{y}_i^2\}$ as the test RESET2.

The RESET test was not designed to have power against a specific alternative hypothesis. However, although the alternative is vague, the choice of test statistic is motivated by a suspected departure from the maintained hypothesis in some particular direction. It tests "regression directions" (Davidson and MacKinnon, 1985, 1987); e.g. for RESET2, the only direction in which the null is false is that represented by $\widehat{y}_i^2$. The RESET is not designed to test against any misspecification which affects higher moments of the error terms than the mean.

[3] This requires that the errors are normally distributed and that the regressors are either fixed in random samples or independent of $\boldsymbol{u}$; more generally, $s\mathrm{F} \sim \chi^2(s)$, asymptotically.

[4] Thursby (1989, p. 222) discusses two-tailed F tests, finding that "F test specification error tests such as the RESET ... are generally most powerful when constituted as one-tail tests. Even when the two-tail test is most powerful, the power is very low on average".

3 The hypergeometric function, ${}_pF_q$

A function $y = f(x)$ is *transcendental* if it may *not* be transformed into a polynomial involving the two variables, x and y, the highest powers of x and y both being greater than unity; and the hypergeometric series is an example of a *higher transcendental* function. Erdélyi, Magnus, Oberhettinger and Tricomi (1953) provide the classic, comprehensive reference for special functions: of particular use are chapters 1 (Gamma function), 2 (Gaussian hypergeometric), 4 (generalized hypergeometric) and 6 (Kummer function). Gradshteyn and Ryzhik (1980, pp. 1039-1059) provide identities and relations for the Gaussian and generalized hypergeometrics and the Kummer function. For alternative coverage of the same material, see Rainville (1960, chs. 4, 5 and 7), Lebedev (1972, chs. 1 and 9), Oberhettinger (1972) for the Gaussian hypergeometric, Slater (1972) for the confluent hypergeometric, Wang and Guo (1989, chs. 4 and 6), Andrews, Askey and Roy (1999) and, in brief, Bell (1968, ch. 9). A less technical approach is found in Sneddon (1980, ch. 2). Jahnke, Emde and Lösch (1960, pp. 4-16) and Lebedev (1972, ch. 1) are useful introductions to the Gamma function; a more detailed treatment is Wang and Guo (1989, ch. 3). Abadir (1999, 2003), Lawford (2001), and Abadir and Rockinger (2003) each cover hypergeometrics with particular relevance to the problems discussed in this paper.

The hypergeometric functions can be generalized in many directions: examples include the very general MacRobert's *E-function* and Meijer's *G-function*; *basic* hypergeometric series; and multiple-argument hypergeometrics (e.g. Appell's hypergeometric functions), as well as forming the basis for computation of parabolic cylinder functions, classes of spherical harmonic functions, etc. Complex and matrix-valued parameters and arguments are all possible. Further study of these generalizations is beyond the scope of this paper. In our applications, we focus upon the hypergeometric series ${}_pF_q$, with real, scalar parameters and real, finite argument, but note the extensions as possible directions for future research.[5]

3.1 The importance of special functions

Special functions have been shown to occur naturally in a variety of theoretical and applied settings in mathematics (e.g. see the survey by Barnard,

[5] Matrix argument hypergeometrics are interesting from a methodological point of view, but application and evaluation may be very difficult indeed, since the general properties of these classes of functions have not yet been worked out with much completeness, e.g. apart from the ${}_0F_0(\boldsymbol{Z})$ and ${}_1F_0(a;\boldsymbol{Z})$, other general explicit results are unknown.

1999; and recent issues of the *Journal of Computational and Applied Mathematics;* theoretical results are often based upon special combinations of parameters), engineering (Kacimov and Obnosov, 2001; Leamy, Noor, and Wasfy, 2001; and Liu, Han and Lam, 2001), environmental science (Kalla and Al-Zamil, 1997), operations research (Halpert, Lengyel, and Pach, 2000), mathematical physics (Ruijsenaars, 1999; and Borodin and Olshanski, 2000; and Seiberg-Witten supersymmetric theory), multivariate statistics (e.g. Möttönen et al., 1998; and Butler and Wood, 2002a, 2002b, 2002c, and references therein); the theory of random matrices (Edelman, Kostlan, and Shub, 1994); and the mathematics of risk (Wang, 2001); and continue to play varied and important roles in many kinds of investigation.

The identification of hypergeometric functions in a particular situation can often facilitate simplified analysis of a problem, since the general properties of many of these functions have been widely established in the mathematical literature. Functional identities and relations between hypergeometrics of different order (e.g. confluences), as well as power series, integral and continued fraction representations, argument transformations, and asymptotic expansions that may be useful for analysis or efficient numerical calculation of the series, can potentially be identified and applied.[6] Derivation of closed-form analytical results in terms of ${}_pF_q$'s instead of ad hoc power series may be of interest in itself; but ideally, this technique will motivate generalizations (from a univariate to a multivariate setting, perhaps).[7]

The generalized hypergeometric function ${}_pF_q$, and both extensions and special cases thereof, have played an important role in derivation of exact finite-sample results in statistics and econometrics since the 1960's (e.g. see the excellent review article by Phillips, 1983). Research by Basmann, Bergstrom, Kabe, Mariano, Phillips, Richardson, Sargan, and Sawa, among others, has established expressions for exact density functions of, e.g., OLS, 2SLS, GCL and LIML estimators for simple structural models; and for testing criteria in a variety of contexts; connections are often seen here between

[6] A simple illustration in econometrics is given by Kleiber (2001). He shows that the relative efficiency of OLS with respect to GLS, in a linear regression model (with a constant term, and long-memory errors) tends to unity as the long-memory parameter d approaches the boundary of the stationary region ($d = 0.5$). The confluence between the ${}_2F_1$ (which arises in the autocorrelation function of a stationary, causal ARFIMA process with distinct roots in the AR polynomial) and the binomial ${}_1F_0$ is useful in simplifying the problem.

[7] An example where the unifying power of special functions is hinted at rather than known arises in the study of cellular automata: "it is my guess that in the end it will in fact turn out to be possible to get a formula for any nested pattern in terms of suitably generalized hypergeometric functions, or perhaps other functions that are direct generalizations of ones used in traditional mathematics." (Wolfram, 2002, p. 612).

the hypergeometrics and integral transforms, or the (non-central) Wishart distribution. Expressing pdf's in terms of hypergeometrics can facilitate analytical study of moments, bias, and distribution functions (e.g. Owen, 1976); with corresponding analysis of existence conditions, and behaviour as certain parameters tend to revealing limits. New applications continue to be discovered, e.g. Hahn and Kuersteiner (2002) give the first two moments of the limiting distributions of 2SLS estimators of a simple simultaneous equations model, with weak instruments and asymptotically vanishing identification, in terms of the ${}_1F_1$; and also Sakalauskas and Šukauskaitė (1996). Recently, infinite series of zonal polynomials, which are closely related to matrix-argument hypergeometrics, have been used to derive closed-form exact results related to ratios of quadratic forms in normal variates, and have provided important insights in the field of analytic distribution theory; see especially Hillier (2001) and Forchini (2002).

The potential of hypergeometrics as a technical tool in unit-root econometrics has begun to be realized. In a series of papers in the 1990's, Abadir (e.g. Abadir, 1995, and references therein) details links of hypergeometrics with unit root test statistics, and highlights issues of convergence and numerical evaluation of the series. Abadir (1993b) derives a closed-form (integral-free) expression for the non-standard *limiting* distribution of the normalized autocorrelation coefficient, given a Gaussian AR(1) with unit root, in terms of nested infinite sums of convergent parabolic cylinder functions: the cdf may be evaluated following truncation of these series, and is easily programmable and highly accurate – for instance, using expansions for the confluent hypergeometric functions derived by Abadir (1993c).[8] In the same framework, Abadir (1995) derives the limiting cdf of the t-statistic for testing for a random walk. An alternative expression for the cdf is given by Dietrich (2001), who uses a property of Liebnitz series to bound the overall approximation error due to series truncation; the cdf formulae in Abadir (1995) and Dietrich (2001) may be manipulated analytically, e.g. in derivation of the limiting pdf, by termwise differentiation.

Abadir (1993a) derives a high-order closed form analytical approximation for the *finite sample* bias of the MLE of the autoregressive parameter in a nonstationary AR(1), also in terms of nested infinite sums involving parabolic cylinder functions. Exact formulae for density functions and moment formulae give rise to series related to (possibly matrix-argument)

[8] Parabolic cylinder functions may also be expressed in terms of Kummer's function, the ${}_1F_1$, although this representation will not necessarily be the most efficient from the point of view of computational generation; e.g. see Abadir (1999).

hypergeometric functions in many areas; however, these are currently intractable in many simple frameworks (e.g. multivariate generalization of Abadir, 1993a); or may be very difficult to implement for numerical evaluation.[9] It may then be necessary to rely upon approximations even when the exact formulae are available, in which case known results in terms of hypergeometrics may sometimes form the basis for heuristic or other approximations (e.g. see Abadir, Hadri and Tzavalis, 1999, and Lawford and Stamatogiannis, 2002, for approximate bias formulae for the MLE in a purely nonstationary VAR(1)).

van Garderen and Shah (2002) derive the exact minimum variance unbiased estimator of the percentage impact of a dummy variable on the level of the dependent variable in a semilog regression equation with normal disturbances; also, its variance, and the exact minimum variance unbiased estimator of the variance.[10] Exact closed-form expressions are given in terms of the ${}_0F_1$, and the ${}_0F_1$ terms are usefully interpreted as a "bias correction for parameter uncertainty" (ibid., p. 151). It is argued that the confluence between the ${}_0F_1$ and ${}_0F_0$ ($\lim_{m\to\infty} {}_0F_1(m; ma) = {}_0F_0(a)$) holds approximately, for m reasonably large. This leads to a simple, accurate and computationally convenient approximation for the exact minimum variance unbiased estimator of the variance, the benefits of which are illustrated in an application to teacher earnings. Likewise, the confluences for the general ${}_pF_q$ can potentially be used to derive such approximations in other settings.

Namba (2002) derives explicit formulae for the exact predictive mean squared error of various biased estimators in a linear regression model (e.g. Stein-rule, minimum MSE), when relevant regressors are omitted; this may be written in terms of nested infinite sums of a scalar argument ${}_2F_1$, whereupon some analytic properties are of use, e.g. convergence and termination of the series expansion. Various characteristic functions (c.f.'s) may be

[9] This is illustrated in a recent Bayesian study by Chao and Phillips (2002), who investigate the behaviour of posterior distributions under the Jeffreys prior in a simultaneous equation model (general limited information setup with $n+1$ endogenous variables). They give a marginal posterior density in terms of the matrix argument ${}_1F_1$, which has an infinite series representation in terms of zonal polynomials. They note the drawback that the density, in this form (ibid., pp. 260-261) "...does not easily lend itself to numerical evaluation, especially in the case where the number of endogenous variables n is greater than two. One difficulty arises because no general formula is known for the zonal polynomials...in the case where $n > 2$, so numerical calculations of the coefficients in the zonal polynomials themselves are also needed." The series may also be very slow to converge; hence, these problems will make exact numerical computation difficult in practice.

[10] General theoretical results are given by van Garderen (2001), based upon a Laplace inversion technique for deriving unbiased predictors in exponential families.

usefully expressed in terms of confluent hypergeometric functions (${}_1F_1$ and Tricomi's function), with complex argument, e.g. Phillips (1982) for the (non-)central F distribution; and Abadir and Magdalinos (2002) for doubly-truncated continuous distributions.[11]

3.2 Useful results for the Gamma function: $\Gamma(m)$

An understanding of the properties of the *Gamma function* $\Gamma(m)$ is a prerequisite for the study of many other special functions. We use the results of this section in implementing both analytic continuation for the Gaussian hypergeometric and asymptotics for the Kummer function. The Gamma function is defined by the *Gauss product*

$$\Gamma(m) \equiv \lim_{n\to\infty}\left[\frac{n!n^m}{m(m+1)(m+2)\cdots(m+n)}\right];\ m \in \mathbb{C}\backslash\mathbb{Z}_{0,-}, \tag{4}$$

which is analytic over the entire complex plane except for simple poles at the nonpositive integers and zero. For $m \in \mathbb{R}_+$, (4) is analogous to the more familiar integral transform definition

$$\Gamma(m) = \int_0^\infty x^{m-1}\exp(-x)\,\mathrm{d}x, \tag{5}$$

which is known as *Euler's second integral*. The single-valued function $\Gamma(m)$ is meromorphic, since it has simple poles only at $\mathbb{Z}_{0,-}$; and is the reciprocal of an entire function, and therefore has no zeros. It satisfies the fundamental functional relation $\Gamma(m+1) = m\Gamma(m)$, where $\Gamma(m+1) = m!$ for $m \in \mathbb{Z}_+$; and the normalization $\Gamma(1) = 1$ is made. Finally, we give an asymptotic result for $\Gamma(m)$: *Stirling's formula*, which states that

$$\Gamma(m+1) \sim (2\pi m)^{1/2}\, m^m \exp(-m)\,,\ \text{as } m \to \infty.$$

When $m \in \mathbb{Z}_+$ and m is large, this is known as *Stirling's factorial approximation*; see Patin (1989) for a concise proof.[12]

[11] Alternative derivations of doubly-truncated c.f.'s are possible, for specific densities, e.g. representation of the c.f. of the doubly-truncated $\mathrm{N}(\mu, \sigma^2)$ as a definite integral, followed by evaluation using substitution. A hypergeometric-based approach here seems to result in some degree of trade-off between generality and simplicity.

[12] $\Gamma(m+1) \geq \sqrt{2\pi m}\, m^m \exp(-m)$ is an *asymptotic to equality* inequality, which holds for $m > 0$. The r.h.s. is in error by 5%, 1% and 0.1% as $m \approx 1.61$, 8.29 and 83.29.

3.2.1 $\Gamma(-m)$ defined for $m \in \mathbb{R}_+ \backslash \mathbb{Z}_+$

Lebedev (1972, p. 4) gives the reflection formula $\Gamma(z)\Gamma(1-z) = \pi(\sin \pi z)^{-1}$, for $z \notin \mathbb{Z}$. Following the transformation $z = m+1$ and using the addition rule $\sin(\theta + \phi) = \sin\theta\cos\phi + \cos\theta\sin\phi$, we find that (for $m \notin \mathbb{Z}$)

$$\Gamma(-m) = \frac{-\pi}{m\Gamma(m)\sin \pi m}. \tag{6}$$

Equations (5) and (6) together define $\Gamma(m)$ everywhere on the real line, except for the poles. This extension has not been incorporated into the GAUSS for Windows NT/95 *gamma*$(\cdot)$ routine, as of version 3.2.32. A different solution is proposed by Forrey (1997, pp. 89-91).

3.3 The generalized hypergeometric function: ${}_pF_q$

Define *Pochhammer's symbol* $(\alpha)_n$ as

$$(\alpha)_n \equiv \alpha(\alpha+1)\cdots(\alpha+n-1) = \Gamma(\alpha+n)\Gamma(\alpha)^{-1};\ n \in \mathbb{Z}_+, \tag{7}$$

with $(\alpha)_0 \equiv 1$ (for $\alpha \neq 0$), and where $(\alpha)_n$ is finite even when $\Gamma(\alpha+n)$ and $\Gamma(\alpha)$ are not analytic;[13] for instance, $-3 = (-3)_1 = \Gamma(-2)\Gamma(-3)^{-1}$. The shifted factorial (7) is an immediate generalization of the elementary factorial, since $n! = (1)_n$. Using (7), we write

$${}_pF_q(a_1,\ldots,a_p;c_1,\ldots,c_q;z) \equiv \sum_{n=0}^{\infty} \frac{(a_1)_n\cdots(a_p)_n}{(c_1)_n\cdots(c_q)_n} \times \frac{z^n}{n!}, \tag{8}$$

as a function of parameters a_k and c_j, and argument z; the prefix and suffix denote the number of numerator and denominator parameters respectively. As shorthand for (8), we use ${}_pF_q(\boldsymbol{a}';\boldsymbol{c}';z)$, where $\boldsymbol{a}$ and $\boldsymbol{c}$ are $p\times 1$ and $q \times 1$ vectors. When it is desired to achieve notational economy by indicating the number of numerator and denominator parameters but not specifying them, we write ${}_pF_q$. It is permissible for either p or q, or both, to be zero. We use the notation ${}_p\boldsymbol{F}_q(\boldsymbol{a}';\boldsymbol{c}';\boldsymbol{z})$ to represent a $\eta \times 1$ vector (where $\boldsymbol{z} = \{z_j\}$ is also $\eta \times 1$) with representative element $\{{}_pF_q(\boldsymbol{a}';\boldsymbol{c}';z_j)\}$.

If any numerator parameter a_k in (8) is zero or a negative integer, the ${}_pF_q$ series will terminate after $(1-a_k)$ terms, the resulting sum being a polynomial of degree $(-a_k)$ in z, and convergence is not an issue. If some

[13] Oldham and Spanier (1974, p. 17) provide the general expression $\Gamma(-v)\Gamma(-w)^{-1} = (-1)^{w-v} w!(v!)^{-1}$, for $(v,w) \in \mathbb{Z}_+^2$.

denominator parameters are $c_j \in \mathbb{Z}_-$, there must exist a corresponding numerator parameter $a_k \in \mathbb{Z}_{0,-}$, with $a_k \geq \max\{c_j : c_j \in \mathbb{Z}_-\}$. This ensures that no zero factors (simple poles) appear in the denominators of the series. Note that no c_j may equal zero. Also, ${}_pF_q(\boldsymbol{a}'; \boldsymbol{c}'; 0) = 1$ follows directly from (8). If $a_k = c_j \neq 0$, for some j, k, so that a_k and c_j *coalesce*, then the corresponding terms in the power series cancel, and ${}_pF_q$ reduces to ${}_{p-1}F_{q-1}$ (e.g. ${}_1F_1(a; a; x) = {}_0F_0(x)$, which is the simplest form of confluence relation). Multiplication (but not division) of parameters is commutative.

Hypergeometric series are convergent within certain domains of their arguments, and an application of a *ratio test* (e.g. Widder, 1989, pp. 288-289) to the power series (8) shows the following:

- If $p < q + 1$, the series is an entire function of z, and converges absolutely when $|z| < \infty$.
- If $p = q + 1$, the series converges for $|z| < 1$ and diverges for $|z| > 1$. Define $\psi = \sum_{j=1}^{q}(c_j - a_j) - a_{q+1}$. The series (8) is absolutely convergent on the unit circle $|z| = 1$ when $\mathrm{Re}(\psi) > 0$, and conditionally convergent for $|z| \neq 1$ when $-1 < \mathrm{Re}(\psi) \leq 0$. When z lies outside of (or, for numerical reasons, within but close to) the boundary of convergence, *analytic continuation* is necessary, with transformation of arguments to lie in the convergence domain. Various transformation formulae are available in the literature, but analytic equivalence does not equal computational equivalence, and different formulae can have very different numerical properties, which raises issues of efficient generation. The hypergeometrics ${}_{p+1}F_p$ are generally more difficult to generate than those for which $p < q + 1$, and especially as p increases.
- If $p > q + 1$, the series diverges for $z \neq 0$. Series of this type are often interpreted as asymptotic series for $z \to 0$.

3.4 The exponential and binomial functions: ${}_0F_0$ and ${}_1F_0$

The most elementary instance of the ${}_pF_q$ is the case for which no numerator or denominator parameters are present. The resulting ${}_0F_0$ is the exponential

$$ {}_0F_0(z) \equiv \sum_{n=0}^{\infty} \frac{z^n}{n!} \equiv \exp(z). \tag{9} $$

In the case of one numerator parameter and no denominator parameters, we obtain the binomial expansion

$$ {}_1F_0\,(a;-z) \equiv \sum_{n=0}^{\infty} \frac{(a)_n\,(-z)^n}{n!} \equiv (1+z)^{-a}\,. \tag{10} $$

We make use of the r.h.s. of these expressions when generating the exponential and binomial; this is far more efficient than use of the series expansions.

3.5 Kummer's function: ${}_1F_1$

When $p = q = 1$, we have *Kummer's confluent (degenerate) hypergeometric function*, for all $a, z \in \mathbb{R}$ and $c \neq 0$:[14]

$$ {}_1F_1\,(a;c;z) \equiv \sum_{n=0}^{\infty} \frac{(a)_n}{(c)_n} \times \frac{z^n}{n!}. \tag{11} $$

If $c \in \mathbb{Z}_-$, it is required that $a - c \in \mathbb{Z}_+$. It satisfies an important identity, *Kummer's Transform*, which constitutes a reflection formula for the ${}_1F_1$:

$$ {}_1F_1\,(a;c;z) \equiv \exp(z)\,{}_1F_1(c-a;c;-z). \tag{12} $$

For some special cases of the ${}_1F_1$, including representations of Hermite polynomials and cylinder functions, see Lebedev (1972, pp. 271-274).

Given dependence upon the argument, the ${}_1F_1$ may or may not exhibit zeros or optima depending upon the values of the two parameters. The ${}_1F_1$ is analytically well-behaved: except when the series terminates (degenerate case) it belongs to an exponential class of functions, and is rapidly convergent for all finite z — the series representation then provides a practical method for calculating ${}_1F_1$. For large values of the argument, numerical values may be calculated most efficiently via the asymptotic representations in Abadir (1999, p. 298), which derive from relations of the ${}_1F_1$ with Tricomi's function, following asymptotic expansion of the latter:

$$ {}_1F_1\,(a;c;z) \sim \begin{cases} \Gamma(c)\,\Gamma(c-a)^{-1}\,|z|^{-a}\,, & z \to -\infty \\ \Gamma(c)\,\Gamma(a)^{-1}\,z^{a-c}\exp(z)\,, & z \to +\infty. \end{cases} \tag{13} $$

Following investigation, we do not suggest use of asymptotics in an automated procedure involving generation of the ${}_1F_1$, since the accuracy of (13) can be shown to vary in a complicated manner with a, c, and z (the unusual "angel's wings" problem discussed in the Appendix).

[14] Alternative notation for Kummer's function in the literature includes $\Phi\,(a;c;z)$, $M\,(a,c,z)$, and $\overset{\infty}{u}\,(a,b,z)$. Our notation clarifies the relationship between the ${}_1F_1$ and the generalized hypergeometric ${}_pF_q$, and permits unambiguous determination of a, c.

3.6 The Gaussian hypergeometric series: ${}_2F_1$

The function ${}_2F_1$ occurs widely in applied problems, largely due to the fact that it is a solution to a certain Fuschian differential equation.[15] The ${}_2F_1$ has a rather complicated structure, and various limiting cases (e.g. the behaviour of ${}_2F_1$ as $z \to 1$ can be classified according to a, b, and c; see Ponnusamy, 1997; and Temme, 2002, for large parameter asymptotics). When the ${}_2F_1$ is represented by a non-terminating infinite series, the radius of convergence is unity. However, for $|z| \geq 1$, analytic continuation is required. Use of argument transformations for convergence of the ${}_2F_1$ is discussed by Forrey (1997), who proposes transformation to the interval $[0, 1/2]$, leading to *rapid* convergence of the resulting series. In this paper, we combine results from Forrey (1997); and from Erdélyi et al (1953, pp. 105-107, eqns. 7, 23, 33) and Lebedev (1972, p. 249) [E and F below, respectively], to give the following transformations (where w is the transformed argument, and $S \equiv {}_2F_1\left(a, b; c; z\right)$).[16]

A $z \in (-\infty, -1) \to w = (1-z)^{-1} \in \left(0, \frac{1}{2}\right)$

$$\begin{aligned} S &= (1-z)^{-a} A_1 \times {}_2F_1\left(a, c-b; a-b+1; (1-z)^{-1}\right) \\ &\quad + (1-z)^{-b} A_2 \times {}_2F_1\left(b, c-a; b-a+1; (1-z)^{-1}\right). \end{aligned}$$

B $z \in [-1, 0) \to w = z\,(z-1)^{-1} \in \left(0, \frac{1}{2}\right]$

Pfaff-Kummer Transform

$$S = (1-z)^{-a}\ {}_2F_1\left(a, c-b; c; z\,(z-1)^{-1}\right).$$

C $z \in \left[0, \frac{1}{2}\right] \to w = z \in \left[0, \frac{1}{2}\right]$

No transformation needed

[15] It is called the Gaussian hypergeometric series following Gauss' examination of this infinite series in his 1812 thesis: *Disquisitiones generales circa seriem infinitam.*

[16] This method improves upon the similar approach proposed by Lawford (2001, pp. 55-56) in two ways: Firstly, the transformed argument lies in $[0, 1/2]$ rather than $[0, 1]$, which increases the speed of convergence. Secondly, Forrey (1997) proposes an approach for dealing with pathological limiting cases, based on finite-difference techniques. We do not implement this here, since we suppose that such parameter combinations will only be encountered on subsets of Lebesgue measure zero (or "very small" subsets) in (a, b, c)–space, and would be too complicated to code for a small expected return in applications/simulations. However, such limiting cases do represent generic solutions to various problems in mathematics.

or *Euler's Transform*

$$S = (1-z)^{c-a-b}\ {}_2F_1\left(c-a, c-b; c; z\right).$$

D $z \in \left(\frac{1}{2}, 1\right] \rightarrow w = 1 - z \in \left[0, \frac{1}{2}\right)$

$$\begin{aligned} S &= A_3 \times {}_2F_1\left(a, b; a+b-c+1; 1-z\right) + (1-z)^{(c-a-b)} \\ &\quad \times A_4 \times {}_2F_1\left(c-a, c-b; c-a-b+1; 1-z\right). \end{aligned}$$

E $z \in (1, 2] \rightarrow w = 1 - z^{-1} \in \left(0, \frac{1}{2}\right]$

$$\begin{aligned} &S = z^{-a} A_3 \times {}_2F_1\left(a, a-c+1; a+b-c+1; 1-z^{-1}\right) \\ &+ z^{a-c}\left(1-z\right)^{c-a-b} A_4 \times {}_2F_1\left(c-a, 1-a; c-a-b+1; 1-z^{-1}\right), \end{aligned}$$

where $z > 1$ ensures that both z^{-a} and z^{a-c} are real for arbitrary a,b and c; however, $(1-z)^{c-a-b}$ may take complex values.

F $z \in (2, \infty) \rightarrow w = z^{-1} \in \left(0, \frac{1}{2}\right)$

$$\begin{aligned} S &= (-z)^{-a} A_1 \times {}_2F_1\left(a, a-c+1; a-b+1; z^{-1}\right) \\ &\quad + (-z)^{-b} A_2 \times {}_2F_1\left(b, b-c+1; b-a+1; z^{-1}\right), \end{aligned}$$

where both $(-z)^{-a}$ and $(-z)^{-b}$ may be complex; and $A_1, \ldots, A_4$ are defined as follows:

$$\begin{aligned} A_1 &= \Gamma(c)\Gamma(b-a)\left\{\Gamma(b)\Gamma(c-a)\right\}^{-1}; \\ A_2 &= \Gamma(c)\Gamma(a-b)\left\{\Gamma(a)\Gamma(c-b)\right\}^{-1}; \\ A_3 &= \Gamma(c)\Gamma(c-a-b)\left\{\Gamma(c-a)\Gamma(c-b)\right\}^{-1}; \\ A_4 &= \Gamma(c)\Gamma(a+b-c)\left\{\Gamma(a)\Gamma(b)\right\}^{-1}. \end{aligned}$$

1. In cases (A) and (F), problems arise when $a - b \in \mathbb{Z}$; in cases (D) and (E), problems arise when $c - a - b \in \mathbb{Z}$; essentially, each of the terms is infinite when taken individually, although their sum remains finite. Some inaccuracies may arise when parameters are close to these cases.

2. Given the following conditions: $b-a > 0$, $a-b > 0$, $c-a-b > 0$, $a+b-c > 0$ (respectively), we can write A_1, A_2, A_3, A_4 as ${}_2F_1\left(a, c-b; c; 1\right)$, ${}_2F_1\left(c-a, b; c; 1\right)$, ${}_2F_1\left(a, b; c; 1\right)$ and ${}_2F_1\left(c-a, c-b; c; 1\right)$.[17]

[17] See Rainville (1960, pp. 48-49) and Bell (1968, pp. 212-213) for proofs of the *Gaussian summation formula* ${}_2F_1\left(a, b; c; 1\right) = \Gamma(c)\Gamma(c-a-b)\left\{\Gamma(c-a)\Gamma(c-b)\right\}^{-1}$.

3.7 (Vector) linear updating

Use of the series expansions, with explicit calculation of Pochhammer terms, (e.g. by recursion), may result in considerable inaccuracies when any of the parameters becomes large. Since a finite number of terms is sufficient for precision to any finite number of decimal places, it is more reasonable to employ a linear updating formula, (Abadir, 1999, pp. 326-327), which relates the successive terms t_{j+1} and t_j in ${}_pF_q = \sum_{j=0}^{\infty} t_j$ as follows:

Scalar update $({}_pF_q)$

$$t_{j+1} \equiv t_j \frac{(a_1+j)\cdots(a_p+j)}{(c_1+j)\cdots(c_q+j)}\left(\frac{z}{j+1}\right);\ t_0 = 1;\ j \in \mathbb{Z}_{0,+}. \qquad (14)$$

Terms are computed until a term becomes zero to the required degree of precision, the remaining terms being truncated. The series associated with $p < q+1$ converge very rapidly, with few terms being required to achieve a high level of precision.

Vectors ${}_0\boldsymbol{F}_0$ and ${}_1\boldsymbol{F}_0$ may be generated using the representations (9) and (10) respectively, for all arguments, i.e. the $n \times 1$ vectors $\{{}_0F_0\}$ and $\{{}_1F_0\}$ may be calculated for all n elements simultaneously (we assume throughout that the parameters $\boldsymbol{a}$ and $\boldsymbol{c}$ are the same for all observations). Matters become more complicated for higher-order hypergeometrics, which require linear updating. It is unwise to consider each element in turn: this will be a computationally intensive procedure, even in the case of few observations, and is impossibly time consuming for high-order hypergeometrics. We present a simple "conservative" solution here in the context of a ${}_pF_q$ series, where asymptotics are used; the approach generalizes without difficulty to evaluation of hypergeometrics which require analytic continuation, and is implemented in the numerical examples, and in the simulations.

Consider an $n \times 1$ vector of real arguments $\boldsymbol{z}$, where $\boldsymbol{z} = \{z_j\}$. We sort the elements of $\boldsymbol{z}$ in ascending order and enter these into an $n \times 1$ vector $\boldsymbol{\psi}$. Note that this operation is valid, since we are not dealing with time series data. Define $S(\boldsymbol{A})$ as the set containing the same elements as $\boldsymbol{A}$, and $\overline{S}(\boldsymbol{A})=\{|x| : x \in S(\boldsymbol{A})\}$. We then partition $\boldsymbol{\psi}$ as $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', \boldsymbol{\psi}_3')'$, where $S(\boldsymbol{\psi}_1) = \{z_j : z_j \leq \tau_1\}$, $S(\boldsymbol{\psi}_2) = \{z_j : \tau_1 < z_j < \tau_2\}$, and $S(\boldsymbol{\psi}_3) = \{z_j : \tau_2 \leq z_j\}$. The vectors $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$, and $\boldsymbol{\psi}_3$ contain n_1, n_2 and n_3 arguments, such that $n_1, n_2, n_3 \geq 0$ and $n_1 + n_2 + n_3 = n$. The constants $\tau_1 < 0, \tau_2 > 0$ determine the "cut-off" points at which we consider asymptotics (if available) to be valid. We define the vector analogue of (14) as

Vector update $({}_pF_q)$

$$\boldsymbol{t}_{j+1} = \frac{(a_1+j)\cdots(a_p+j)}{(c_1+j)\cdots(c_q+j)(j+1)}(\boldsymbol{t}_j \odot \boldsymbol{\psi}_2);\ \boldsymbol{t}_0 = \boldsymbol{\imath}_{n_2};\ j \in \mathbb{Z}_{0,+},$$

where $\boldsymbol{t}_j \equiv \{t_j\}$, and $\odot$ is the Hadamard product. We see that convergence is determined by finding the element $\gamma = \arg\max\left(\overline{S}(\boldsymbol{\psi}_2)\right)$, and then examining the corresponding element of the linear update vector, $t_j(\gamma)$. Computationally, we may update $\boldsymbol{\psi}_2$ as a vector, while convergence of $\{t_j(\gamma)\}_{j=1}^{\infty}$ ensures convergence of the whole of $\boldsymbol{\psi}_2$. The precision of element $t_j(\gamma)$ was at least 9 decimal places in applications.

1. Importantly, it is necessary to re-order the vector of hypergeometric terms once the linear update has been completed.

2. This technique is applied to all the hypergeometrics used in the paper that require linear updating (e.g. ${}_0F_1$, ${}_0F_2$ and ${}_1F_2$). The procedure must be implemented several times when argument transformations are used (e.g. Kummer's Transform for the ${}_1F_1$, and Euler's and other transforms in the case of the ${}_2F_1$).

3. We illustrate the gains of the vector linear update by generating a Kummer function for multiple arguments, without asymptotics. We generated two random vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, of size $n = 100$ and $n = 1,000$, with elements drawn from the uniform $U(-5,5)$. We then computed the $n \times 1$ vector ${}_1\boldsymbol{F}_1(2;3;\boldsymbol{z}_j)$ for $j = 1,2$. The term-by-term linear update needed at least 0.11 seconds and 0.99 seconds respectively. The vector linear update required no more than 0.01 seconds and 0.06 seconds respectively to produce the resorted vector of results. In terms of a simulation study, using 10,000 replications, 100 observations, and generating 10 hypergeometrics in each of 35 optimization iterations, this corresponds to a saving of 4 days CPU time.

4 The hypergeometric test

We now use the theory detailed above to construct a new test for additive omitted nonlinearity, given that the correct class of functional forms of the omission is known; we then discuss details of its implementation. The assumption of correct class was a useful simplification for the present work, but will be relaxed in future studies. We propose the $\mathrm{H}({}_pF_q)$ test as follows. The null model is identical to that under the RESET procedure, (1):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}.$$

The hypergeometric series is then used to replace the auxiliary regression of the RESET (2) with the more general form

$$\boldsymbol{y} = h\left(\boldsymbol{\theta}; \boldsymbol{X}, \widehat{\boldsymbol{y}}\right) + \boldsymbol{u}, \tag{15}$$

where $h(\cdot)$ is a vector-valued function

$$h\left(\boldsymbol{\theta}; \boldsymbol{X}, \widehat{\boldsymbol{y}}\right) = b_0\left(\boldsymbol{X}\boldsymbol{\beta} + {}_p\boldsymbol{F}_q\left(\boldsymbol{a}'; \boldsymbol{c}'; \boldsymbol{m} + b_1\widehat{\boldsymbol{y}}\right)\right),$$

such that $h : \mathbb{R}^{n\times(k+1)} \times \mathbb{R}^{r\times 1} \to \mathbb{R}^{n\times 1}$, when $p \leq q$, and $h : \mathbb{R}^{n\times(k+1)} \times \mathbb{R}^{r\times 1} \to \mathbb{C}^{n\times 1}$, when $p = q+1$; and ${}_p\boldsymbol{F}_q$ is an $n \times 1$ vector with jth element $\{{}_pF_q\left(\boldsymbol{a}'; \boldsymbol{c}'; m + b_1\widehat{y}_j\right)\}$, $\boldsymbol{m} = \{m\}$, and $\widehat{\boldsymbol{y}} = \{\widehat{y}_j\}$. The unknown scalar parameters b_0, b_1 and m are included for scaling purposes.

The parameters $\boldsymbol{\theta} = \mathrm{vec}\left(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{\beta}, m, b_0, b_1\right)$ form an $(r+1)\times 1$ vector, and are estimated using nonlinear least squares: $\min_{\boldsymbol{\theta}} \Omega$, with objective surface

$$\Omega = \left\| \boldsymbol{y} - h\left(\boldsymbol{\theta}; \boldsymbol{X}, \widehat{\boldsymbol{y}}\right)\right\|_2,$$

where $\|\boldsymbol{w}\|_2 = \sqrt{\langle \boldsymbol{w}, \boldsymbol{w}\rangle} = \sqrt{\boldsymbol{w}'\overline{\boldsymbol{w}}}$ is the Euclidean norm, and $\overline{\boldsymbol{w}}$ is the complex conjugate of $\boldsymbol{w}$.[18] Other methods of estimation are discussed by Abadir and Rockinger (2003). When $p = q + 1$ (see remark 3 below), the objective surface is modified by inclusion of a Lagrangean penalty, and the optimization problem becomes $\min_{\boldsymbol{\theta}} \{\Omega + \lambda p(\boldsymbol{u})\}$, with

$$p(\boldsymbol{u}) = \mathrm{Im}(\boldsymbol{u})' \,\mathrm{Im}(\boldsymbol{u})\,; \text{ and } \lambda \in \mathbb{R}_+, \text{ where } \lambda \text{ is large.}$$

[18] Care must be exercised when evaluating such norms in GAUSS 3.2, e.g. $\boldsymbol{w} = (2+i, 3-2i)'$ should give $\boldsymbol{w}'\overline{\boldsymbol{w}} = 18$. However, the seemingly obvious w'conj(w) returns $8+8i$. The correct command is sumc(w.*conj(w)).

We used $\lambda = 100$ in the simulations. We test the null hypothesis $H_0 : b_1 = 0$ (no additive nonlinearity) against the alternative $H_1 : b_1 \neq 0$ by estimating (1) and (15), and then using the pseudo-F likelihood ratio statistic

$$\mathrm{F}_H = \left(\frac{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{\boldsymbol{v}^{\star\prime}\boldsymbol{v}^{\star}} - 1 \right) (n - r) \,.$$

We estimate r parameters under the alternative (m is fixed; see remark 5 below), and impose a single restriction under the null. We refer to the distribution $\mathrm{F}(1, n-r)$ for reference purposes, when considering the empirical size of F_H. Since the true asymptotic and small-sample distributions are unknown, we assess the null using right-hand simulated critical values of $\widetilde{\mathrm{F}}_H$ under the null; it is rejected at a $100\alpha\%$ level of significance when $\mathrm{F}_H > \widetilde{\mathrm{F}}_H^{\alpha}$. Under the null, $h\left(\mathrm{vec}\left(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{\beta}, m, b_0, 0\right); \boldsymbol{X}, \widehat{\boldsymbol{y}}\right)$ reduces to a constant, which essentially shifts the identification problem discussed by Hansen (1996) onto b_0. The numerically minimized sum of squared residuals under the alternative is denoted by $\boldsymbol{v}^{\star\prime}\boldsymbol{v}^{\star}$. The parameter m is chosen to prevent any numerator/denominator parameters from disappearing.

Practical considerations

1. The above formulation of the $\mathrm{H}({}_pF_q)$ test is stylized, and allows for various possible refinements. It was chosen to illustrate the potential gains of using hypergeometrics in a destructive testing framework: it combines notational simplicity and computational tractability, and allows comparison with the RESET test, given that the correct class of functional forms of the omission is known. Modifications might include consideration of a greater range of ${}_pF_q$'s; more effective numerical calculation of hypergeometrics (wrt accuracy, speed); improved numerical optimization; as well as introduction of additional parameters within $h(\cdot)$, for greater flexibility.

2. If p and/or q are too large, we will encounter numerical problems analogous to micronumerosity, i.e. estimation of too many parameters, given too small a sample size. In addition, general, efficient evaluation routines are not available for high-order hypergeometrics; e.g. analytic continuation of ${}_3F_2$, ${}_4F_3$, etc. will be very complicated, as are asymptotics for general ${}_pF_q$. We therefore focus attention on ${}_0F_0$, ${}_0F_1$, ${}_0F_2$, ${}_1F_1$, ${}_1F_2$ ($p \leq q$), and ${}_1F_0$, ${}_2F_1$ ($p = q + 1$). Despite the efficient algorithm presented for the ${}_2F_1$, we were unable to incorporate it into our simulation study; essentially, reliable estimation

of the ${}_2F_1$ parameters was very slow, and some convergence problems were encountered. This does not preclude use of the ${}_2F_1$ for testing, given an improved estimation routine; and we illustrate the usefulness of this function for constructive modelling, in the Appendix, where a grid-search is used across some of the ${}_2F_1$ parameters.

3. When $p \leq q$, the ${}_pF_q$ series converges absolutely for all $|z| < \infty$, and the hypergeometric cannot be complex-valued if both parameters and argument are real. However, when evaluating the ${}_{p+1}F_p$ outside of $|z| < 1$, with use of a closed-form (${}_1F_0$), or analytic continuation (${}_2F_1$), the hypergeometric function *may* be complex-valued; e.g. ${}_1F_0\left(-\frac{1}{2}; 5\right) = +\sqrt{-4} = 2i$, and ${}_2F_1\left(\frac{1}{3}, \frac{5}{6}; \frac{2}{3}; 5\right) \approx 0.1872 - 0.4840i$ [in both cases, the unmodified ${}_pF_q$ series will diverge]. Implementation of the test procedure, without correction for complex-valued hypergeometrics, was shown to severely reduce the power of the test for the ${}_1F_0$; power results given an artificial modification of the objective surface Ω for the ${}_1F_0$ (see Lawford, 2001) were poor, and mild convergence problems arose; neither case is reported here.

4. We considered two earlier formulations of the test, based upon

$$h(\boldsymbol{\theta}; \boldsymbol{X}, \widehat{\boldsymbol{y}}) = \boldsymbol{X}\boldsymbol{\beta} + b_0\, {}_p\boldsymbol{F}_q\left(\boldsymbol{a}'; \boldsymbol{c}'; \boldsymbol{m} + b_1\widehat{\boldsymbol{y}}\right),$$

for both $H_{01} : b_0 = 0$ and $H_{02} : b_1 = 0$, against the two-sided alternatives. Use of H_{01} creates a difficulty, since $\boldsymbol{a}$, $\boldsymbol{c}$, m and b_1 are not identified under the null. When a hypothesis on some parameters causes other parameters to disappear from the model, then nonstandard asymptotics arise. We do not wish to fix the hypergeometric parameters in advance, since this would reduce both flexibility and power. While Hansen (1996) provides a theoretical framework for inference in such cases, his approach seems too complicated to apply here (a multidimensional grid-search, with the parameter space replaced by a discrete approximation, would be very expensive). We considered two simpler solutions: $H_{02} : b_1 = 0$, which was seen to lead to poor power in simulations; and the preferred formulation used in this paper. Although (15) does not reduce to (1) under the null $H_0 : b_1 = 0$, the parameter b_0 is a constant under the null; and by the Frisch-Waugh decomposition theorem, this parameter has no effect upon the distribution of the residual sum of squares.

5. In a series of simulation studies, we assessed a variety of values for m, both floating and fixed. The case $\overline{m} = -1$ (where the bar denotes a fixed parameter) gave the best power results among the variants and is reported below; this was seen to be a good choice for all the $\mathrm{H}({}_pF_q)$ tests which *require* $m \neq 0$ (e.g. for $\mathrm{H}({}_1F_0)$, we need $m \notin \{0, 1\}$ and $a \neq 0$, so that no parameters disappear).

6. The function ${}_0F_1$ converges absolutely for all real finite arguments and is thus real everywhere. It is the simplest hypergeometric that requires generation by the linear updating algorithm. The remaining hypergeometrics of interest pose potential identification problems, since the search routine may be unable to distinguish between two numerator or two denominator parameters, when these are close in magnitude. As a result, the optimization routine may loop forever. We solve this by an ordering of the parameters in the hypergeometric function. Whenever two numerator and/or denominator parameters are present, we define $a_2 \equiv a_1 + |a_2'|$ and/or $c_2 \equiv c_1 + |c_2'|$. For instance, we fit ${}_1F_2\left(a; c_1, c_1 + |c_2'|\,; m + b_1\widehat{y}_j\right)$.

7. The $\mathrm{H}({}_pF_q)$ test is based upon a numerical optimization. For this reason, the choice of starting values is of some importance and requires particular attention. We noticed a number of interesting features during the optimizations, which enabled us to considerably reduce the CPU time needed for the simulations. The starting values for the parameters of $h(\cdot)$ are denoted by a superscript $\dagger$ throughout, and we use superscript $\star$ to refer to the numerically optimized parameters.

   [**Parameter signs**] The *sign* of any numerator or denominator parameters, or of b_1, did not change from the sign of the starting values (e.g. if $a^\dagger > 0$, then $a^\star > 0$ always resulted). The optimization routine was unable to find optima associated with some optimized parameters of the opposite sign to their starting value. To facilitate correct optimization, it was necessary to consider both positive and negative values for a_k, $(k = 1, 2, \ldots, p)$, and for b_1. We also noted a lack of convergence when $a^\dagger$ was chosen to be a negative integer: the starting parameter did not change over the iterations in some cases. We suggest that $a^\dagger \in \mathbb{Z}_{-,0}$ should *not* be used in practical applications.

   [**Singularities**] Denominator parameters pose an additional problem when any $c^\dagger < 0$ is chosen. The series expansion exhibits regular singularities at $c \in \mathbb{Z}_{-,0}$. If $c^\dagger \in [\varrho, \varrho + 1]$, where $\varrho \in \mathbb{Z}_-$, then $c^\star \in [\varrho, \varrho + 1]$ always results, i.e. the denominator parameter is "trapped" within the two contiguous singularities. In this paper, we always chose $c^\dagger > 0$; in future work, it would be of some interest to consider negative starting values for $c^\dagger$, e.g. $c^\dagger \in \{(-5,-4), \ldots, (-1,0), \mathbb{R}_+\}$.

[**Invariance properties**] The optimizations were invariant across a large range of $\alpha^\dagger$ and $\beta^\dagger$. The parameter b_0 was able to change sign from that of its starting value. In fact, $b_0^\star > 0$ always resulted, due in part to the particular choice of DGP, and optimization results were unaffected by the choice of $b_0^\dagger$.

[**Kummer function**] When either p or q is zero, the size of any numerator or denominator starting values is apparently of little importance. However, when both p and q are non-zero, interaction effects seem to arise. For instance, the Kummer series ${}_1F_1$ satisfies the Kummer transformation (12). Hence, we may wish to choose $a^\dagger$ and $c^\dagger$ such that both signs of $c^\dagger - a^\dagger$ are considered (e.g. if $a^\dagger = \pm 2.5$ and $c^\dagger = 3.5$, we have violent results, of an exponential nature). We suggest that $c^\dagger$ is selected to be small in absolute value relative to $a^\dagger$ in such cases.

If we were interested in considering every combination of *signs* of starting values, to be more certain of finding global rather than local optima, then we would require 2^r separate optimizations (or 2^{r+1} if m is estimated) for each replication, e.g. ${}_1F_2$ would need 128 starts, which is not possible given available computational power. From the above points, we suggest reducing the number of starts by always using $\alpha^\dagger = \widehat{\alpha}$, $\beta^\dagger = \widehat{\beta}$ and $b_0^\dagger = 1$. Also, $c_k^\dagger > 0$ for each denominator parameter. We assess all combinations of signs of any numerator parameters a_k, and of b_1. Thus, the required number of starts is reduced to $2^{p+1} << 2^r$ (e.g. ${}_1F_2$ requires 4 starts). For instance, when $\mathrm{H}({}_1F_0)$ is implemented, we use starting parameters $(\alpha^\dagger, b_1^\dagger) = (\pm 2.5, \pm 1)$, where each combination of signs is considered, in addition to $\alpha^\dagger = \widehat{\alpha}$, $\beta^\dagger = \widehat{\beta}$, $b_0^\dagger = 1$, and $\overline{m} = -1$. The minimum value of the optimized objective function is chosen from the four optimizations.[19]

No other restrictions were placed upon the parameters and only minor convergence difficulties were encountered (in several $\mathrm{H}({}_1F_1)$ replications). It is unwise to attempt any updating of starting values, since

[19] Analogous problems arise in the implementation of artificial neural networks (ANN's), which often apply the *multistart method.* This uses multiple starting points; of those that converge, the "best" are chosen or combined. Another method of dealing local versus global ANN optima is to perturb the point in question and check whether the optimization continues to return to the same point. Multiple starts are also used in semi-nonparametric modelling, e.g. see Brunner (1992) for a discussion.

some regions of convergence may always lead to local optima. Ideally, we would like to assess many random starting values, although this is currently feasible only in a single implementation of the test, or in a modelling rather than testing context. The starting values that we used are given in Table 1, with approximate CPU times needed to simulate critical values with $10,000$ replications and $N = 100$.

8. We also assessed replacement of $\widehat{y}_j$ with X_j, which gave critical values that were identical to those derived when $\widehat{y}_j$ was used; this is unsurprising, since $\widehat{y}_j$ is an affine transformation of X_j.

9. Different optimization methods have varying degrees of success in different applications and a great deal of experimentation was required. An iterative algorithm may search forever unless a termination criterion is specified. Hence, it was important to experiment with tolerance to assess the sensitivity of the results to changes in stopping rules. Since we are formulating our tests in terms of SSR, finding an exact minimum is much less important than ensuring that further decreases in SSR are marginal. The routine may be halted at this stage. Since our choice of tolerance is not always satisfied, due to a very flat objective surface near to the optimum, we specify a maximum number of iterations of 35, after which the optimization terminates. It was necessary to verify in each simulation of $\mathrm{H}({}_pF_q)$ that the value of the objective surface would change little given additional iterations.

10. Table 2 summarizes the generation methods for each of the $\mathrm{H}({}_pF_q)$ tests: (1) is a linear update required? (2) do we impose a Lagrangean penalty against complex-valued hypergeometrics? (3) do we order numerator and/or denominator parameters? (4) are asymptotics considered? [see Section 3.5] (5) are argument transformations applied?

5 Monte Carlo study

In this section, we assess the small-sample distributional properties of the hypergeometric $\mathrm{H}({}_pF_q)$ test statistic; and also its ability to detect additive omitted nonlinearity, when the class of functional forms of the omission is known. We present the main results of an extensive simulation study that compares $\mathrm{H}({}_pF_q)$ against the RESETs. All simulations were run on a Pentium 3, 450MHz machine, with 128MB of RAM, using GAUSS 3.2 and the CO optimization module, under Microsoft Windows 98; and all reported CPU times in the paper refer to this specification. At a conservative estimate, $7-8,000$ hours of CPU time were used in the development and investigation of the properties of the hypergeometric tests.

5.1 Setup

We use the simple data generating process (DGP)

$$y_j = \alpha_1 + \alpha_2 X_j + \alpha_3\, g\,(X_j) + \varepsilon_j, \tag{16}$$

where α_1, α_2 and α_3 are scalar parameters; $\varepsilon_j \sim \mathrm{N}(0,1)$; $j = 1, \ldots, N$, where N is the sample size; and $g\,(\cdot)$ is some nonlinear function, which will be chosen to correspond to a particular *class* of hypergeometric functions. The functions $g\,(\cdot)$ must be scaled carefully, in order to show the relative power differences between the competing tests. We give size and power results for samples of $N = 25$, 50, 100 observations, and we use $10,000$ replications in all cases. Since we condition on $\{X_j\}_{j=1}^{N}$, we only report results for $X_j \sim \mathrm{N}(0,5)$. The generated set of $\{X_j\}_{j=1}^{N_1}$ is equal to the first N_1 observations generated for a larger set $\{X_j\}_{j=1}^{N_2}$, $N_2 > N_1$. We use (16) to construct the pairs $\{(y_j, X_j)\}_{j=1}^{N}$, given parameter values $(\alpha_1, \alpha_2, \alpha_3)' = (2, 0.5, 1)'$, and with $\{X_j\}_{j=1}^{N}$ generated once. We found no qualitative changes when different parameter values were chosen, and some of our Figures display output from these other studies. We report results for simulated quantiles to 2 d.p. but caution against interpreting these beyond 1 d.p.; however, the power results are quite robust against changes in the number of replications.

5.2 Results on null distributions

We obtain empirical null distributions for the $\mathrm{H}({}_pF_q)$ tests using distribution sampling, and then calculate finite-sample tail quantiles, which we use in our assessment of power. We assess the size of each $\mathrm{H}({}_pF_q)$ test, since they will be implemented separately: in each instance, we test for a given omitted nonlinear term by fitting the corresponding hypergeometric function (as opposed to parameters) within (16). Since the limiting distribution of our test statistic is unknown, reference is made to the $\mathrm{F}(1, N-r)$, computed using the GAUSS *cdffc* routine; and to the $\chi^2\,(1)$, in Figure 1.

Table 3 reports simulated tail quantiles for each of the hypergeometric test statistics; and we use 5% critical values in implementing the tests. The Table required approximately 2 weeks of CPU time to calculate. We report in parentheses the quantiles for the corresponding $\mathrm{F}(1, N-r)$ distributions.

We see comparable results for $\mathrm{H}({}_0F_0)$, $\mathrm{H}({}_1F_0)$, $\mathrm{H}({}_0F_1)$, $\mathrm{H}({}_0F_2)$ and $\mathrm{H}({}_1F_2)$: 5% quantiles tend to a figure of roughly 4, as N increases. The test statistics follow positively-skewed distributions and the distribution of $\widetilde{\mathrm{F}}_H$ appears to be reasonably well approximated, to 1 d.p., by $\mathrm{F}(1, N-r)$ in terms of 10% and 5% quantiles, as N becomes larger. However, this observation does not hold for the 25% and 1% quantiles. Results for $\mathrm{H}({}_1F_1)$ are markedly different; a very small number of $\mathrm{H}({}_1F_1)$ replications failed to converge, in which case we drop the offending replication and continue.[20] This result for the $\mathrm{H}({}_1F_1)$ test is as yet unexplained. We calculate the ratio $\varsigma(\widetilde{\mathrm{F}}_H)$ of 1% to 5% critical values for the pseudo-F hypergeometric tests. Since this is larger than the corresponding ratio $\varsigma(\mathrm{F})$ for the $\mathrm{F}(1, N-r)$, it implies that the $\mathrm{H}({}_pF_q)$ distribution has thicker tails, which become more pronounced when the variance of the $\mathrm{H}({}_pF_q)$ test statistic is smaller than that for $\mathrm{F}(1, N-r)$.

In Table 4, we report the simulated means and variances of the $\mathrm{H}({}_pF_q)$ test statistics. We calculate the first two central moments of the F distribution (Spanos, 1986, p. 113)

$$\mathrm{E}\left(\mathrm{F}\left(m,n\right)\right) = \frac{n}{n-2},\ \left(n>2\right),$$

and

$$\mathrm{var}\left(\mathrm{F}\left(m,n\right)\right) = \frac{2n^2\left(m+n-2\right)}{m\left(n-2\right)^2\left(n-4\right)},\ \left(n>4\right),$$

and compare them with the simulated $\mathrm{E}(\widetilde{\mathrm{F}}_H)$ and $\mathrm{var}(\widetilde{\mathrm{F}}_H)$. This again shows the different shape taken by the $\mathrm{H}({}_1F_1)$ test statistic. As the sample size increases, the other $\mathrm{H}({}_pF_q)$ tests have comparable mean and variance. The simulated variances were often seen to fall as the sample size increased.

In Figure 1, we present box plots of the $\mathrm{H}({}_pF_q)$ test statistics, for $N = 100$. We remove the largest 5% of the sample, since some extreme observations obscured the results. The horizontal lines corresponding to each "case" represent (from bottom to top of the Figure): minimum, 25th percentile (75% quantile), median, 75th percentile and maxmimum of the *truncated* sample. Hence, the topmost horizontal line represents the 5% critical value. The boxes represent interquartile ranges. We generate $10,000$ observations from the $\chi^2(1)$ and include these for comparison. Clearly, all of the densities are highly (positively) skewed. Results are alike for all of the

[20] We discovered that the quantity $\widetilde{\mathrm{F}}_H < -0.05$ was a reasonable indicator that the numerical optimization had failed. It was found that $\widetilde{\mathrm{F}}_H \in [-0.05, 0]$ sometimes resulted from numerical inaccuracies and indicated that the OLS solution had in fact been reached.

$\mathrm{H}({}_pF_q)$ tests except the $\mathrm{H}({}_1F_1)$, which is reflected in the simulated quantiles of Table 3. Whenever the median is unclear, it is closer to the 25th than the 75th percentile. In Table 5, we assess the normality of the estimated parameters using the Jarque-Bera test. Normality is strongly rejected at all conventional small-sample or asymptotic significance levels.[21]

5.3 Results on power

We concentrate upon the following questions here: Firstly, are the small-sample powers of the $\mathrm{H}({}_pF_q)$ tests an improvement over those of the RESET tests? Secondly, is the performance of each hypergeometric test good against a variety of nonlinearity in that class? Tables 6–11 present power results (using simulated critical values from Table 3 for the $\mathrm{H}({}_pF_q)$ test and exact critical values for the RESETs) against a range of nonlinear omissions. When one of the tests outperforms the other two, we report the power in bold type. We see that the RESET2 often has higher power than the RESET, although examination of individual replications shows that RESET can sometimes correctly identify an omission when RESET2 does not. We expect the RESETs to perform well whenever the omission $g(\cdot)$ can be well approximated by some finite, low-order polynomial.

Significant small-sample power gains are seen for the $\mathrm{H}({}_pF_q)$ test over the RESETs for a variety of $g(\cdot)$, in some cases, this represents a three-fold increase. These gains are generally reduced as N increases except for the $\mathrm{H}({}_1F_2)$, where results perhaps indicate a failure to converge for the $\mathrm{H}({}_1F_2)$ test for small N. In Figures 2–4, we plot fitted values for the RESETs and appropriate $\mathrm{H}({}_pF_q)$ tests, against $g(X_j) = \exp(0.3Xj)$, $g(X_j) = (1 + 0.2X_j)^2$, and $g(X_j) = {}_0F_1(2; 0.7X_j)$; using different DGP parameters in each case. The $\mathrm{H}({}_pF_q)$ fits are visibly different from those of the RESETs. A mild inconsistency is seen for one of the ${}_0F_2$ omissions, although this is much less severe than the loss suffered by the RESET2. When the $\mathrm{H}({}_pF_q)$ test is outperformed by the RESETs, the relative loss is small. However, the $\mathrm{H}({}_1F_1)$ test is seen to behave differently in Table 10, and performs rather worse relative to the RESETs than the other $\mathrm{H}({}_pF_q)$ tests; although it does still have higher power that the RESET in some instances.

[21] Lawford (2001, p. 115, Tables 3.8-3.10) examined six non-normal error pdfs: chi-squared $\chi^2(2)$, lognormal $\mathrm{LN}(2,1)$, Fisher's $\mathrm{F}(4,8)$, Student's $\mathrm{t}(5)$, Cauchy $C(0,1)$ and uniform $U(0,1)$, all transformed to be $\mathrm{iid}(0,1)$, for $\mathrm{H}({}_0F_0)$, $\mathrm{H}({}_1F_0)$ and $\mathrm{H}({}_0F_1)$; these tables require several weeks of CPU time for calculation, and show that 5% critical values change considerably under non-normal errors, except for $\mathrm{t}(5)$ and $U(0,1)$.

6 Concluding remarks

This paper has developed a new class of tests for omitted nonlinearity, based on using the generalized hypergeometric function ${}_pF_q$ in a novel manner, and assuming that the correct class of functional forms of the omission is known. We have constructed a workable procedure for implementing these tests, and offer practical solutions to a number of unexpected theoretical and numerical difficulties; and also give evidence of the speed and accuracy of our method of ${}_pF_q$ generation. We have designed a simple Monte Carlo experiment which illustrates our findings. The approach appears promising, and the power performance of the hypergeometric tests is good relative to the RESETs. However, this is a very new field, and we freely acknowledge that many questions are still unanswered. We end with a short discussion of directions for future research, which are under investigation by the author.

The first interesting topic is the further assessment of the limiting distribution of the $\mathrm{H}({}_pF_q)$ tests under the null, with a view to answering the following question: Is the limiting distribution the same for different p and q (which would indicate numerical problems in implementing $\mathrm{H}({}_1F_1)$), or not (in which case the quantiles for $\mathrm{H}({}_1F_1)$ demand a theoretical explanation)? It is possible that ${}_{p+j}F_{q+j}$ may have similar quantiles for various j, (or perhaps a pattern of increase with j), but not across different p and q. A response surface to describe how the quantiles vary with sample size N, p and q in ${}_pF_q$ would be very useful; this would solve some issues related to how a practitioner should choose critical values.

Secondly, it is important to consider a parsimonious hypergeometric test for omitted nonlinearity given that the correct class of functional forms of the omission is *unknown*. While Abadir (1999) proposes a sequential general-to-specific approach, based upon the confluences between ${}_pF_q$'s of different order, (which are potentially testable, although no application of this yet exists in the literature), we believe that a model-selection-based procedure merits consideration, à la Akaike and Schwarz, (see also Sin and White, 1996). Theoretical justification for the penalty function would be required. Some rigorous theoretical foundations of this extension, and of the current paper (e.g. consistency of test) remain to be developed.

A third strand of research would be to investigate new constructive modelling procedures, as opposed to the destructive test developed in this paper. There is a strong indication that constructive modelling using ${}_pF_q$'s is viable, and can offer significant benefits over fully (non)parametric methods, when applied to datasets that are not large; in addition, the estimated hypergeometric parameters can reveal both functional form and shape parameters.

7 Appendix: testing the hypergeometric code

We cross-checked our code against published results in the applied mathematics and statistics literature wherever possible, and report some findings on speed and accuracy below. All numerical scalar ${}_pF_q$ computations were verified using Mathematica 4.1. Some additional guidelines for computational generation of the hypergeometric functions are given by Abadir (1999, pp. 326-330). Other interesting methods are available for particular ${}_pF_q$'s, e.g. Gautschi (2002) [uses Gauss-Jacobi quadrature to approximate integral representations associated with the ${}_1F_1$ and ${}_2F_1$, for real parameters and complex argument: a major limitation of this approach is that the integral forms are admitted only when some parameters are restricted by certain inequalities: $b > a > 0$ for the ${}_1F_1$; and $c > b > 0$ for the ${}_2F_1$].

[7.1] Richardson and Wu (1971) compare the OLS and 2SLS estimators of structural coefficients in a certain simultaneous equation model. Use of our Kummer routine reveals several errors in calculations based upon the ${}_1F_1$: in comparison of the biases of the estimators (ibid., p. 977, Table I; and their equation (3.1)), the reported value of 0.65, for entry $N - K_1 = 100$, $K_2 = 5$, $\mu^2 = 2$, is incorrect; the actual value is

$$\frac{{}_1F_1\,(1.5; 2.5; 1)}{{}_1F_1\,(49; 50; 1)} = 0.70674369.$$

In comparison of MSE's (ibid., p. 979, Table II; and their equation (4.1)), entries $N-K_1 = 50$, $K_2 = 3$, $\mu^2 = 5$, and $\left|\overline{\beta}\right| = 3, 4, 5, 10$ are incorrect: 0.39, 0.34, 0.33, 0.32 should be replaced by 0.3433, 0.3286, 0.3218, 0.3127. The ${}_1F_1$ appears in Richardson and Wu (1970), in expressions for the bias and MSE of the OLS estimator of the slope coefficient in an errors-in-variables model. We see that entry $n = 3$, $\tau = 15$ in (ibid., p. 743) Tables A-1 and A-2a should be corrected as follows: $0.0211 \to 0.0228$; and $0.1056 \to 0.1138$.

[7.2] We test the linear update by using *Saalschütz's Theorem* :

$$\begin{aligned} C &\equiv {}_3F_2\,(a, b, c; d; e; 1) \\ &= \frac{\Gamma\,(d-a+|c|)\,\Gamma\,(d-b+|c|)\,\Gamma\,(d)\,\Gamma\,(d-a-b)}{\Gamma\,(d-a)\,\Gamma\,(d-b)\,\Gamma\,(d+|c|)\,\Gamma\,(d-a-b+|c|)} \equiv D, \end{aligned}$$

where $d + e = a + b + c + 1$ and $c \in \mathbb{Z}_-$. Define $\mathcal{A} = \{1, 2, \ldots, n\}$ and $\mathcal{B} = \left\{\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n+1}\right\}$, where $n \in \mathbb{Z}_+$ and choose parameters as follows:

$c \in -\mathcal{A}$; $(a, b) \in \mathcal{A} \times \mathcal{A}$ subject to $a \leq b$; $e \in \mathcal{B}$ and $d = a + b + c + 1 - e$. Then, $C(a, b, c, d, e)$ terminates after n terms; the restriction on (a, b) prevents us from counting permutations (e.g. $(1, 2)$ and $(2, 1)$); and d and e are non-integer (and $d \neq e \neq 0$) so that no cancellation occurs between numerator and denominator parameters, and $C(a, b, c, d, e)$ has no singularities. Choosing the parameters in this way gives $\frac{1}{2}n^3(n+1)$ combinations. We calculated $C(a, b, c, d, e)$ and $D(a, b, c, d, e)$ and found:

n	$\frac{1}{2}n^3(n+1)$	$\max\limits_{a,b,c,d,e}\{\lvert C-D\rvert\}$
6	756	1.1504×10^{-10}
7	1372	2.1239×10^{-9}
8	2304	2.8729×10^{-8}
9	3645	4.2885×10^{-7}
10	5500	5.8254×10^{-6}

The reduction in accuracy occurs since $\inf \mathcal{B} = 0$ (i.e. the denominator of the series approaches a singularity as n increases). The highest-order ${}_pF_q$ that we used in testing the linear update was taken from Krupnikov and Kölbig (1997, p. 94, eqn. (7.7.2.10)): $G \equiv \frac{8}{9}\,{}_6F_5\left(\frac{1}{4}, \frac{1}{4}, \frac{3}{4}, \frac{3}{4}, 1, \frac{3}{2}; \frac{1}{2}, \frac{5}{4}, \frac{5}{4}, \frac{7}{4}, \frac{7}{4}; 1\right)$ ≈ 0.91596555, which evaluates the *Catalan constant* G correctly to 7 d.p., after 888 terms.

[7.3] In order to illustrate the benefits of our method for evaluating the ${}_2F_1$, we computed both sides of the following relation, taken from Oberhettinger (1972, p. 556, eqn. (15.1.14)):

$$ {}_2F_1\left(a, a + \frac{1}{2}; 2a; z\right) = 2^{2a-1}(1-z)^{-\frac{1}{2}}\left[1 + (1-z)^{\frac{1}{2}}\right]^{1-2a}, \; z \in \mathbb{R} \setminus \{1\}, \tag{17} $$

which is complex-valued for $z > 1$. We set $a = \frac{1}{3}$ and define the 2000×1 vector of arguments $\boldsymbol{z} = \{-10, -9.99, \ldots, 9.99, 10\} \setminus \{1\}$. Generation of the individual elements ${}_2F_1\left(\frac{1}{3}, \frac{5}{6}; \frac{2}{3}; (\boldsymbol{z})_j\right)$ required 3.25 seconds. The vector generation routine required no more than 0.05 seconds on the same machine; we display results below, and plot this function in Figure 5. A simple linear update would face problems of very slow convergence close to $|z| < 1$.

z	w	Update	Terms	${}_2F_1$ routine and equation (17)	Terms all series
-10	0.09091	n/a	n/a	0.38965037	14
-5	0.16667	n/a	n/a	0.48959014	19
-1.1	0.47619	n/a	n/a	0.73827487	45
-1.01	0.49751	n/a	n/a	0.75138430	48
-0.9999	0.49998	0.75290868	162, 232	0.75290868	22
-0.999	0.49975	0.75304506	17, 333	0.75304506	22
-0.99	0.49749	0.75441376	1, 837	0.75441376	22
-0.9	0.47368	0.76861431	186	0.76861431	21
0.9	0.1	2.7506396	186	2.7506396	15
0.99	0.01	8.1932127	1, 837	8.1932127	8
0.999	0.001	25.360840	17, 333	25.360840	4
0.9999	0.0001	79.63374**1**	162, 232	79.633742	4
1.01	0.00990	n/a	n/a	0.26407928 $-7.9457917i$	8
1.1	0.09091	n/a	n/a	0.25989290 $-2.5368116i$	15
1.9	0.47368	n/a	n/a	0.23308328 $-0.90144966i$	49
1.99	0.49749	n/a	n/a	0.23082583 $-0.86434872i$	53
2.01	0.49751	n/a	n/a	0.23033855 $-0.85679275i$	59
2.1	0.47619	n/a	n/a	0.22820640 $-0.82540808i$	55
5	0.2	n/a	n/a	0.18719874 $-0.48400513i$	26
10	0.1	n/a	n/a	0.15705026 $-0.35515658i$	18

[7.4] Practical implementation of ${}_1F_1(a;c;z)$ asymptotics leads to what we call the "angel's wings" problem: In Figure 6, we plot $\mu :=$ ratio of asymptotic (13) to linear update (11) [or (12)], against $a-c$, for 2,000 realizations of $(a,c) \sim U(-5,5)^2$, and for $z=50$. Define $\mathcal{Y}(z) := \{\mu\}$, and the approximate probability of attaining no more than a 5% error through use of asymptotics as $p(z) := \text{card}(\mathcal{Y}(z) \cap [0.95, 1.05]) / \text{card}(\mathcal{Y}(z))$ [card≡"cardinality"]. Choice of a constant cut-off at which asymptotics are to be applied is very unwise, since μ clearly depends upon a and c, and the range of error from using the asymptotics may be large (e.g. $\inf \mathcal{Y}(50) \approx 0.22$, $\sup \mathcal{Y}(50) \approx 1.20$; and $p(50) \approx 0.3218$).[22] Since the relationship between μ and a, c, z is unknown, development of a reliable automated procedure will be difficult. Although the accuracy of asymptotics improves as $|z|$ increases, errors are substantial even for large argument (e.g. $p(100) \approx 0.4818$) [see also Figure 7, which plots μ against z, for ${}_1F_1(-1.5;3;z)$].

[7.5] The quantities $\psi = \sum_{k=1}^{q} c_k - \sum_{j=1}^{p} a_j$ and $Q = \prod_{j=1}^{p} |a_j| \left(\prod_{k=1}^{q} |c_k|\right)^{-1}$ have both been suggested as heuristic measures of the "fragility" of convergence of a series ${}_{p+1}F_p$, for given argument z; they measure the "weight" of the numerator parameters relative to the denominator parameters. Following experimentation, we note that increasing ψ tends to reduce the number of linear update terms needed for convergence; however, given ψ, reducing Q also tends to reduce the number of terms needed.

[7.6] When generating the ${}_1F_1$, there is a clear trade-off between using the *exact* linear update, (which will become inefficient as the argument increases), and the asymptotic *approximation*, (which is inaccurate for small argument). When $z < 0$, the terms of the linear update form an alternating series and may lead to numerical inaccuracies for large negative argument (e.g. ${}_1F_1(2;3;-40) \approx 0.051026550$, which is incorrect). We suggest using (12) when $z < 0$ to transform $z \longmapsto -z$, whereupon a linear update may be performed on the transformed function (which gives the correct result: ${}_1F_1(2;3;-40) = 0.001250000$; note that ${}_1F_1(2;3;z) = z^{-2}\{2 + 2e^z(z-1)\}$). We also found that Butler and Wood's (2002a) calibrated ${}_1F_1$ Laplace approximation performs better than the asymptotics for ${}_1F_1(2;3;z)$, across $z \in [-40, 40]$.

[22] Estimates are from 5,000 realizations of (a,c).

[7.7] **Modelling nonlinearity using the ${}_2F_1$:** Although it currently seems too difficult to implement the ${}_2F_1$ in an automated setting, we illustrate the usefulness of the generation routines discussed in this paper, in a constructive modelling problem. We treat an example that was considered by Chen, Lockhart and Stephens (2002), consisting of $n = 107$ observations on distance travelled in kilometres (Y_i) and gasoline used in litres (X_i). A simple Box-Cox transformation is

$$Y(\vartheta) = \begin{cases} \left(Y^{\vartheta} - 1\right)/\vartheta, & \vartheta \in \mathbb{R} \setminus \{0\} \\ \ln Y, & \vartheta = 0, \end{cases} \tag{18}$$

where ϑ is selected – usually by a grid search – so that the linear model $Y_i(\vartheta) = \mu + \beta X_i + \varepsilon_i$, $i = 1, 2, \ldots, n$, is approximately applicable (this will arguably have certain desirable properties). Assuming that the errors are $\mathrm{N}\left(0, \sigma^2\right)$, the maximum likelihood estimate $\widehat{\vartheta}$ of ϑ is found as the value which maximizes the profile log-likelihood. Based upon a grid search across $\vartheta \in \{0.50, 0.51, ..., 2.50\}$, we find $\widehat{\vartheta} \approx 1.47$. The observation that (18) may be written more generally as (Abadir, 1999)

$$Y^1(\vartheta) = (Y - 1)\ {}_2F_1(1 - \vartheta, 1; 2; 1 - Y), \ \vartheta \in \mathbb{R},$$

suggests that the Box-Cox transformation may be generalized by replacing (18) with a hypergeometric functional form (either involving a ${}_2F_1$ or, more generally, a ${}_pF_q$; with estimation of some or all parameters). To illustrate, we assess the simple generalization

$$Y^2(\vartheta) = (Y - 1)\ {}_2F_1(1 - \vartheta, \psi; 2; 1 - Y), \ (\vartheta, \psi) \in \mathbb{R}^2.$$

We note that the log-likelihood function must now be modified for a Jacobian based upon the ${}_2F_1$, which is termwise differentiable if it is convergent. Searching in a neighbourhood of the solution found using (18), we estimate $\widehat{\vartheta} \approx 1.46$ and $\widehat{\psi} \approx 1.43$, (with higher log-likelihood), using $Y^2(\vartheta)$. Further development of this technique is underway by the author.

References

1. Abadir, K. M. (1993a): "OLS bias in a nonstationary autoregression," *Econometric Theory*, 9: 81-93.

2. Abadir, K. M. (1993b): "The limiting distribution of the autocorrelation coefficient under a unit root," *Annals of Statistics*, 21: 1058-1070.

3. Abadir, K. M. (1993c): "Expansions for some confluent hypergeometric functions," *Journal of Physics A*, 26: 4059-4066 (corrigendum for printing error, 1993, 26: 7663).

4. Abadir, K. M. (1995): "The limiting distribution of the t ratio under a unit root," *Econometric Theory*, 11: 775-793.

5. Abadir, K. M. (1999): "An introduction to hypergeometric functions for economists," *Econometric Reviews*, 18: 287-330.

6. Abadir, K. M. (2003): *Explicit Distribution Theory for Simple Time Series*, Wiley Series in Probability and Mathematical Statistics, forthcoming.

7. Abadir, K. M., Hadri, K. and Tzavalis, E. (1999): "The influence of VAR dimensions on estimator biases," *Econometrica*, 67: 163-181.

8. Abadir, K. M. and Magdalinos, A. (2002): "The characteristic function from a family of truncated normal distributions," Solution 01.5.2, *Econometric Theory*, 18: 1276-1287.

9. Abadir, K. M. and Magnus, J. R. (2002): "Notation in econometrics: A proposal for a standard," *Econometrics Journal*, 5: 76-90.

10. Abadir, K. M. and Rockinger, M. (2003): "Density functionals, with an option–pricing application," *Econometric Theory*, forthcoming.

11. Abken, P., Madan, D. B. and Ramamurtie, S. (1996): "Estimation of risk-neutral and statistical densities by Hermite polynomial approximation: with an application to Eurodollar futures options," Federal Reserve Bank of Atlanta, mimeo. Abstract and further details at http://www.frbatlanta.org/publica/work_papers/WP965.HTM

12. Al-Saqabi, B. N., Kalla, S. L. and Tuan, V. K. (2003): "Unified probability density function involving a confluent hypergeometric function of two variables," *Applied Mathematics and Computation*, forthcoming.

13. Andrews, G. E., Askey, R. and Roy, R. (1999): *Special Functions*, Cambridge: Cambridge University Press. (Encyclopedia of Mathematics and its Applications, 71)

14. Anscombe, F. J. (1961): "Examination of residuals," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: UOC Press, pp. 1-36.

15. Anscombe, F. J. and Tukey, J. W. (1963): "The examination and analysis of residuals," *Technometrics*, 5: 141-160.

16. Barnard, R. W. (1999): "On applications of hypergeometric functions," *Journal of Computational and Applied Mathematics*, 105: 1-8.

17. Bell, W. W. (1968): *Special Functions for Engineers and Scientists.* London: D. Van Nostrand Limited.

18. Borodin, A. and Olshanski, G. (2000): "Distributions on partitions, point processes, and the hypergeometric kernel," *Communications in Mathematical Physics*, 211: 335-358.

19. Box, G. and Cox, D. (1964): "An analysis of transformations," *Journal of the Royal Statistical Society, Series B*, 26: 211-252 (with discussion).

20. Brunner, A. D. (1992): "Conditional asymmetries in real GNP: A seminonparametric approach," *Journal of Business and Economic Statistics*, 10: 65-72.

21. Butler, R. W. and Wood, A. T. A. (2002a): "Laplace approximations for hypergeometric functions with matrix argument," *Annals of Statistics*, 30: 1155-1177.

22. Butler, R. W. and Wood, A. T. A. (2002b): "Mixture representations of noncentral distributions in multivariate analysis with application to a dimensional CLT," unpublished manuscript.

23. Butler, R. W. and Wood, A. T. A. (2002c): "Approximation of power in multivariate analysis," unpublished manuscript.

24. Chao, J. C. and Phillips, P. C. B. (2002): "Jeffreys prior analysis of the simultaneous equations model in the case with $n+1$ endogenous variables," *Journal of Econometrics*, 111: 251-283.

25. Chen, G., Lockhart, R. A. and Stephens, M. A. (2002): "Box-Cox transformations in linear models: Large sample theory and tests of normality," *Canadian Journal of Statistics*, 30 (with discussion).

26. Davidson, J. (2000): *Econometric Theory*, Oxford: Blackwell.

27. Davidson, R. and MacKinnon, J. G. (1985): "The interpretation of test statistics," *Canadian Journal of Economics*, 18: 38-57.

28. Davidson, R. and MacKinnon, J. G. (1987): "Implicit alternatives and the local power of test statistics," *Econometrica*, 55: 1305-1329.

29. Dietrich, F. K. (2001): "The limiting distribution of the t-ratio for the unit root test in an AR(1)," *Econometrics Journal*, 4: 242-256.

30. Edelman, A., Kostlan, E. and Shub, M. (1994): "How many eigenvalues of a random matrix are real?," *Journal of the American Mathematical Society*, 7: 247-267.

31. Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F. G. (1953): *Higher Transcendental Functions*, volume 1. New York: McGraw-Hill Book Company, Inc. (Bateman Manuscript Project).

32. Forchini, G. (2002): "The exact cumulative distribution function of a ratio of quadratic forms in normal variables, with application to the AR(1) model," *Econometric Theory*, 18: 823-852.

33. Forrey, R. C. (1997): "Computing the hypergeometric function," *Journal of Computational Physics*, 137: 79-100.

34. Gallant, A. R., Rossi, P. E. and Tauchen, G. (1992): "Stock prices and volume," *Review of Financial Studies*, 5: 199-242.

35. Gallant, A. R. and Tauchen, G. (1989): "Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications." *Econometrica*, 57: 1091-1120.

36. Gautschi, W. (2002): "Gauss quadrature approximations to hypergeometric and confluent hypergeometric functions," *Journal of Computational and Applied Mathematics*, 139: 173-187.

37. Giacomini, R., Gottschling, A., Haefke, C. and White, H. (2002): "Hypernormal densities," Discussion Paper in Economics 2002–14, University of California, San Diego. Available from the website http://www.econ.ucsd.edu/papers/2002.html

38. Godfrey, L. G. (1988): *Misspecification Tests in Econometrics*, Cambridge: Cambridge University Press.

39. Gordy, M. B. (1998a): "A generalization of generalized beta distributions," Finance and Economics Discussion Series No. 1998-18, Federal Reserve Board, available from the website http://www.federalreserve.gov/pubs/feds/1998/199818/199818abs.html

40. Gordy, M. B. (1998b): "Computationally convenient distributional assumptions for common-value auctions," *Computational Economics*, 12: 61-78.

41. Gottschling, A., Haefke, C. and White, H. (1999): "Closed form integration of artificial neural networks with some applications to finance," Discussion Paper in Economics 99–24, University of California, San Diego. Available at http://www.econ.ucsd.edu/papers/dp99.html

42. Gradshteyn, I. S. and Ryzhik, I. M. (1980): *Table of Integrals, Series and Products.* London: Academic Press Limited, 5th edition.

43. Granger, C. W. J. and Teräsvirta, T. (1993): *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

44. Hahn, J. and Kuersteiner, G. (2002): "Discontinuities of weak instrument limiting distributions," *Economics Letters*, 75: 325-331.

45. Halpert, A., Lengyel, F. and Pach, J. (2000): "Cellular telephone networks and random maps in hypergraphs," *Discrete Applied Mathematics*, 103: 111-126.

46. Hansen, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64: 413-430.

47. Hillier, G. (2001): "The density of a quadratic form in a vector uniformly distributed on the n-sphere" *Econometric Theory*, 17: 1-28.

48. Jahnke, E., Emde, F. and Lösch, F. (1960): *Tables of Higher Functions.* New York: McGraw-Hill Book Company, Inc.

49. Kacimov, A. R. and Obnosov, Yu. V. (2001): "Analytical solutions by the hodograph method to hydrodynamic problems for oil and gas traps," *Journal of Hydrology*, 254: 33-46.

50. Kalla, S. L. and Al-Zamil, A. (1997): "Solution of a diffusion of dust problem in terms of hypergeometric functions," *Mathematical and Computer Modelling*, 26: 87-93.

51. Kleiber, C. (2001): "Finite sample efficiency of OLS in linear regression models with long-memory disturbances," *Economics Letters*, 72: 131-136.

52. Krupnikov, E. D. and Kölbig, K. S. (1997): "Some special cases of the generalized hypergeometric function ${}_{q+1}F_q$," *Journal of Computational and Applied Mathematics*, 78: 79-95.

53. Kuan, C.-M. and White, H. (1994): "Artificial neural networks: An econometric perspective," *Econometric Reviews*, 13: 1-143.

54. Kumar, C. S. (2002): "Extended generalized Hypergeometric probability distributions," *Statistics and Probability Letters*, 59: 1-7.

55. Lawford, S. (2001): "Improved modelling in finite-sample and nonlinear frameworks," unpublished D.Phil. thesis, University of York.

56. Lawford, S. and Stamatogiannis, M. P. (2002): "The finite-sample effects of VAR dimensions on MLE bias, MLE variance, and minimum MSE estimators: Purely nonstationary case," Discussion Paper in Economics, No. 2002/04, University of York.

57. Leamy, M. J., Noor, A. K. and Wasfy, T. M. (2001): "Sensitivity analysis of bare-wire space tether systems," *Computer Methods in Applied Mechanics and Engineering*, 190: 5495-5503.

58. Lebedev, N. N. (1972): *Special Functions and their Applications.* New York: Dover Publications, Inc.

59. Liu, G. R., Han, X. and Lam, K. Y. (2001): "An integration technique for evaluating confluent hypergeometric functions and its application to functionally graded materials," *Computers and Structures*, 79: 1039-1047.

60. Loeb, P. D. (1976): "Specification error tests and investment functions," *Econometrica*, 44: 185-194.

61. Madan, D. B. and Milne, F. (1994): "Contingent claims valued and hedged by pricing and investing in a basis," *Mathematical Finance*, 4: 223-245.

62. McDonald, J. B. (1984): "Some generalized functions for the size distribution of income," *Econometrica*, 52: 647-663.

63. Möttönen, J., Hettmansperger, T. P., Oja, H. and Tienari, J. (1998): "On the efficiency of affine invariant multivariate rank tests," *Journal of Multivariate Analysis*, 66: 118-132.

64. Namba, A. (2002): "PMSE performance of the biased estimators in a linear regression model when relevant regressors are omitted," *Econometric Theory*, 18: 1086-1098.

65. Oberhettinger, F. (1972): "Hypergeometric functions," in Abramowitz, M. and Stegun, I. A. (eds.), *Handbook of Mathematical Functions*, chapter 15. New York: Dover Publications, Inc.

66. Oldham, K. B. and Spanier, J. (1974): *The Fractional Calculus*. New York: Academic Press.

67. Owen, A. D. (1976): "A proof that both the bias and the mean square error of the two-stage least squares estimator are monotonically non-increasing functions of sample size," *Econometrica*, 44: 409-411.

68. Patin, J. M. (1989): "A very short proof of Stirling's formula," *American Mathematical Monthly*, 96: 41-42.

69. Phillips, P. C. B. (1982): "The true characteristic function of the F distribution," *Biometrika*, 69: 261-264.

70. Phillips, P. C. B. (1983): "Exact small sample theory in the simultaneous equations model," in Griliches, Z. and Intriligator, M. D. (eds.) *Handbook of Econometrics*, volume 1, Amsterdam: North-Holland.

71. Podivinsky, J. M. (1996): "SNP: A program for non-parametric time series analysis," *Economic Journal*, 106: 1825-1829.

72. Ponnusamy, S. (1997): "Close-to-convexity properties of Gaussian hypergeometric functions," *Journal of Computational and Applied Mathematics*, 88: 327-337.

73. Rainville, E. D. (1960): *Special Functions*. New York: Macmillan Publishing Company.

74. Ramsey, J. B. (1969): "Tests for specification error in classical linear least squares regression analysis," *Journal of the Royal Statistical Society, Series B*, 31: 350-371.

75. Ramsey, J. B. (1983): "Diagnostic tests as residual analysis: Perspective and comments," *Econometric Reviews*, 2: 241-248.

76. Ramsey, J. B. and Alexander, A. (1984): "The econometric approach to business cycle analysis reconsidered," *Journal of Macroeconomics*, 6: 347-356.

77. Ramsey, J. B. and Gilbert, R. (1972): "A Monte Carlo study of some small sample properties of tests for specification error," *Journal of the American Statistical Association*, 67: 180-186.

78. Ramsey, J. B. and Schmidt, P. (1976): "Some further results on the use of OLS and BLUS residuals in specification error tests," *Journal of the American Statistical Association*, 71: 389-390.

79. Richardson, D. H. and Wu, D.-M. (1970): "Least squares and grouping method estimators in the errors in variables model," *Journal of the American Statistical Association*, 65: 724-748.

80. Richardson, D. H. and Wu, D.-M. (1971): "A note on the comparison of ordinary and two-stage least squares estimators," *Econometrica*, 39: 973-981.

81. Ruijsenaars, S. N. M. (1999): "A generalized hypergeometric function satisfying four analytic difference equations of Askey-Wilson type," *Communications in Mathematical Physics*, 206: 639-690.

82. Sakalauskas, L. and Šukauskaitė, D. (1996): "Distribution of MLE of the multiple correlation coefficient for data with missing values in a dependent variable," *Statistics and Probability Letters*, 30: 53-59.

83. Sin, C. Y. and White, H. (1996): "Information criteria for selecting possibly misspecified parametric models," *Journal of Econometrics*, 71: 207-225.

84. Slater, L. J. (1972): "Confluent hypergeometric functions," in Abramowitz, M. and Stegun, I. A. (eds.), *Handbook of Mathematical Functions*, chapter 13. New York: Dover Publications, Inc.

85. Sneddon, I. N. (1980): *Special Functions of Mathematical Physics and Chemistry*. London: Longman, 3rd edition.

86. Spanos, A. (1986): *Statistical Foundations of Econometric Modelling*, Cambridge: Cambridge University Press.

87. Temme, N. M. (2002): "Large parameter cases of the Gauss hypergeometric function," *Journal of Computational and Applied Mathematics*, forthcoming.

88. Theil, H. (1965): "The analysis of disturbances in regression analysis," *Journal of the American Statistical Association*, 60: 1067-1079.

89. Theil, H. (1968): "A simplification of the BLUS procedure for analyzing regression disturbances," *Journal of the American Statistical Association*, 63: 242-251.

90. Thursby, J. G. (1989): "A comparison of several specification error tests for a general alternative," *International Economic Review*, 30: 217-230.

91. Thursby, J. G. and Schmidt, P. (1977): "Some properties of tests for specification error in a linear regression model," *Journal of the American Statistical Association*, 72: 635-641.

92. van Garderen, K. J. (2001): "Optimal prediction in loglinear models," *Journal of Econometrics*, 104: 119-140.

93. van Garderen, K. J. and Shah, C. (2002): "Exact interpretation of dummy variables in semilogarithmic equations," *Econometrics Journal*, 5: 149-159.

94. Wang, G. (2001): "A decomposition of the ruin probability for the risk process perturbed by diffusion," *Insurance: Mathematics and Economics*, 28: 49-59.

95. Wang, Z. X. and Guo, D. R. (1989): *Special Functions*. Singapore: World Scientific Publishing.

96. Widder, D. V. (1989): *Advanced Calculus*. New York: Dover Publications, 2nd edition.

97. Wolfram, S. (2002): *A New Kind of Science*, Wolfram Media, Inc.

Table 1: Starting values; $b_0^\dagger = 1$ always chosen

Test	2^{p+1}	Starting values	Time (hrs)
$\mathrm{H}({}_0F_0)$	2	$b_1^\dagger : \pm 1$	3
$\mathrm{H}({}_1F_0)$	4	$(b_1^\dagger, a^\dagger) : (\pm 1, \pm 2.5)$	9
$\mathrm{H}({}_0F_1)$	2	$(b_1^\dagger, c^\dagger) : (\pm 1, 3.5)$	11
$\mathrm{H}({}_0F_2)$	2	$(b_1^\dagger, c_1^\dagger, c_2^\dagger) : (\pm 1, 2.5, 3.5)$	11
$\mathrm{H}({}_1F_1)$	4	$(b_1^\dagger, a^\dagger, c^\dagger) : (\pm 1, \pm 2.5, 0.8)$	60
$\mathrm{H}({}_1F_2)$	4	$(b_1^\dagger, a^\dagger, c_1^\dagger, c_2^\dagger) : (\pm 1, \pm 2.5, 0.8, 1.2)$	30

Table 2: Generation of hypergeometric tests

Test	Update	Penalty	Ordering	Asympts.	Transf.
$\mathrm{H}({}_0F_0)$	no	no	no	no	no
$\mathrm{H}({}_1F_0)$	no	yes	no	no	no
$\mathrm{H}({}_0F_1)$	yes	no	no	no	no
$\mathrm{H}({}_0F_2)$	yes	no	yes	no	no
$\mathrm{H}({}_1F_1)$	yes	no	no	yes	yes
$\mathrm{H}({}_1F_2)$	yes	no	yes	no	no

Table 3: Tail quantiles of $H({}_pF_q)$ test statistics; $F(1, N-r)$ in (.)

Test	N	25%	10%	5%	1%	$\varsigma(\tilde{F}_H)$	$\varsigma(F)$
	25	0.67 (1.40)	1.94 (2.96)	3.00 (4.32)	6.27 (8.02)	2.09	1.85
$H({}_0F_0)$	50	1.15 (1.36)	2.74 (2.82)	4.10 (4.05)	7.33 (7.22)	1.79	1.78
$(r=4)$	100	1.07 (1.34)	2.72 (2.76)	3.87 (3.94)	7.21 (6.91)	1.86	1.75
	25	0.70 (1.40)	1.92 (2.97)	2.99 (4.35)	6.12 (8.10)	2.05	1.86
$H({}_1F_0)$	50	1.05 (1.36)	2.52 (2.82)	3.75 (4.06)	6.86 (7.23)	1.83	1.78
$(r=5)$	100	1.20 (1.34)	2.75 (2.76)	4.02 (3.94)	7.23 (6.91)	1.80	1.75
	25	0.83 (1.40)	2.20 (2.97)	3.44 (4.35)	6.88 (8.10)	2.00	1.86
$H({}_0F_1)$	50	0.95 (1.36)	2.49 (2.82)	3.79 (4.06)	7.01 (7.23)	1.85	1.78
$(r=5)$	100	1.01 (1.34)	2.74 (2.76)	3.98 (3.94)	7.36 (6.91)	1.85	1.75
	25	0.80 (1.41)	2.27 (2.99)	3.59 (4.38)	7.15 (8.18)	1.99	1.87
$H({}_0F_2)$	50	0.87 (1.36)	2.46 (2.82)	3.81 (4.06)	7.07 (7.25)	1.85	1.78
$(r=6)$	100	0.75 (1.34)	2.35 (2.76)	3.67 (3.94)	6.92 (6.91)	1.89	1.75
	25	2.42 (1.41)	4.25 (2.99)	5.67 (4.38)	9.74 (8.18)	1.72	1.87
$H({}_1F_1)$	50	3.14 (1.36)	5.27 (2.82)	6.87 (4.06)	10.49 (7.25)	1.53	1.78
$(r=6)$	100	3.00 (1.34)	5.11 (2.76)	6.71 (3.94)	10.59 (6.91)	1.58	1.75
	25	1.78 (1.41)	3.49 (3.01)	4.84 (4.41)	8.17 (8.29)	1.69	1.88
$H({}_1F_2)$	50	1.17 (1.36)	3.00 (2.83)	4.39 (4.07)	8.07 (7.26)	1.84	1.79
$(r=7)$	100	0.77 (1.34)	2.49 (2.76)	3.83 (3.94)	7.33 (6.92)	1.91	1.75

Table 4: Simulated moments of $H({}_pF_q)$ test statistics

Test	N	$E(\tilde{F}_H)$	E(F)	$var(\tilde{F}_H)$	var(F)
$H({}_0F_0)$ ($r=4$)	25	0.63	1.105	1.80	2.874
	50	0.89	1.045	2.59	2.342
	100	0.84	1.021	2.36	2.154
$H({}_1F_0)$ ($r=5$)	25	0.62	1.111	1.64	2.932
	50	0.82	1.047	2.24	2.351
	100	0.92	1.022	2.35	2.156
$H({}_0F_1)$ ($r=5$)	25	0.72	1.111	2.18	2.932
	50	0.77	1.047	2.29	2.351
	100	0.82	1.022	2.47	2.156
$H({}_0F_2)$ ($r=6$)	25	0.72	1.118	2.24	2.998
	50	0.75	1.048	2.36	2.360
	100	0.71	1.022	2.08	2.157
$H({}_1F_1)$ ($r=6$)	25	1.79	1.118	4.31	2.998
	50	2.29	1.048	5.21	2.360
	100	2.12	1.022	5.27	2.157
$H({}_1F_2)$ ($r=7$)	25	1.29	1.125	3.46	3.074
	50	0.92	1.049	3.09	2.369
	100	0.74	1.022	2.29	2.159

Table 5: Normality of estimated parameters ($N = 100$)

Test	Parameter	SK	KT	JB
H($_0F_0$)	b_0	5.33	34.76	467662.95
	b_1	-2.62	10.49	34854.42
H($_1F_0$)	a	0.0066	0.0035	3741.39
	b_0	61.59	5059.56	1.07x10^{10}
	b_1	-1.15	24.64	197293.23
H($_0F_1$)	c	0.0052	0.0030	3742.53
	b_0	1.40	464.48	8.87x10^{7}
	b_1	-0.00039	0.010	3724.78
H($_0F_2$)	c_1	1.03	4.34	2502.33
	c_2	0.82	4.83	2504.32
	b_0	27.21	2195.46	2.00x10^{9}
	b_1	-0.0013	0.00088	3747.81
H($_1F_1$)	a	0.12	1.35	1156.41
	c	5.76	97.17	3.75x10^{6}
	b_0	0.88	3.73	1506.90
	b_1	2.32	13.55	55298.90
H($_1F_2$)	a	-0.18	0.42	2838.73
	c_1	5.80	46.25	835306.60
	c_2	2.72	16.93	93156.47
	b_0	-1.71	48.38	863092.56
	b_1	-0.34	4.20	794.80

Table 6: Power of $H({}_0F_0)$ test

$g(X_j)$	N	RESET	RESET2	$H({}_0F_0)$
$\exp(0.3X_j)$	25	12	19	**28**
	50	25	36	**44**
	100	78	86	**89**
$0.5\exp(0.4X_j)$	25	13	20	**34**
	50	25	35	**44**
	100	80	84	**89**
$0.8\exp(-0.3X_j)$	25	7	8	**11**
	50	14	**21**	18
	100	92	**96**	93

Table 7: Power of $H({}_1F_0)$ test

$g(X_j)$	N	RESET	RESET2	$H({}_1F_0)$
$(1-0.1X_j)^{-2}$	25	11	16	**33**
	50	18	26	**35**
	100	70	72	**79**
$(1+0.2X_j)^2$	25	8	10	**27**
	50	15	23	**31**
	100	60	77	**80**
$0.1(1+0.5X_j)^3$	25	30	45	**60**
	50	59	71	**78**
	100	**98**	91	96

Table 8: Power of H($_0F_1$) test

$g(X_j)$	N	RESET	RESET2	H($_0F_1$)
$_0F_1(2; 0.7X_j)$	25	9	13	**23**
	50	17	26	**35**
	100	59	73	**77**
$0.4\,{}_0F_1(-1.5; 0.4X_j)$	25	21	31	**43**
	50	36	46	**57**
	100	92	79	**94**
$-0.4\,{}_0F_1(-0.8; 0.2X_j)$	25	11	17	**28**
	50	23	36	**48**
	100	76	88	**90**

Table 9: Power of H($_0F_2$) test

$g(X_j)$	N	RESET	RESET2	H($_0F_2$)
$_0F_2(0.5, 1.5; 0.5X_j)$	25	9	12	**21**
	50	18	27	**36**
	100	67	82	**85**
$2\,{}_0F_2(1, 5; 1.5X_j)$	25	7	10	**16**
	50	14	21	**29**
	100	50	68	**74**
$0.2\,{}_0F_2(-0.5, -2.5; X_j)$	25	96	81	96
	50	**100**	37	86
	100	100	91	100

Table 10: Power of H($_1F_1$) test

$g(X_j)$	N	RESET	RESET2	H($_1F_1$)
$_1F_1(0.2; -0.5; 0.25X_j)$	25	9	**12**	7
	50	14	**21**	8
	100	46	**56**	42
$0.1\,_1F_1(-3; 0.5; 0.3X_j)$	25	7	8	8
	50	12	**19**	14
	100	58	**74**	62
$_1F_1(2; 3; 0.4X_j)$	25	11	17	**21**
	50	21	**31**	25
	100	71	**79**	75

Table 11: Power of H($_1F_2$) test

$g(X_j)$	N	RESET	RESET2	H($_1F_2$)
$0.1\,_1F_2(3; 1, 2; X_j)$	25	17	**27**	25
	50	34	48	48
	100	91	93	**96**
$_1F_2(-1.5; 1, 2; X_j)$	25	6	**8**	4
	50	11	**16**	14
	100	37	54	**59**
$_1F_2(0.5; 1.5, 2.5; -1.5X_j)$	25	6	6	4
	50	9	12	**13**
	100	48	63	**68**

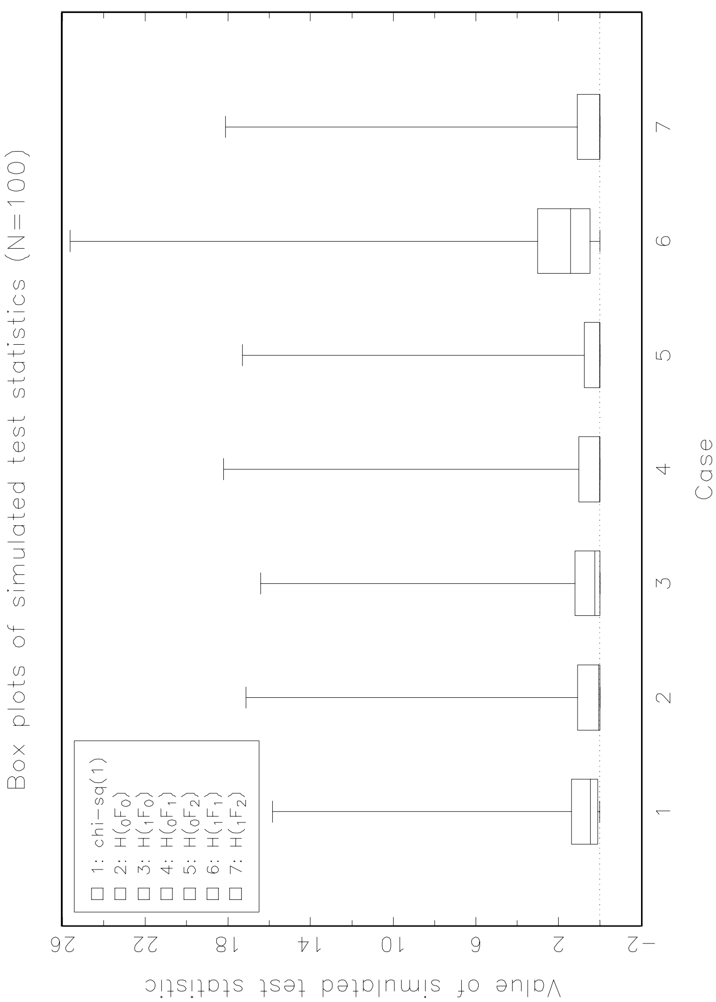

Figure 1: Boxplots of simulated null distributions of H($_pF_q$) tests

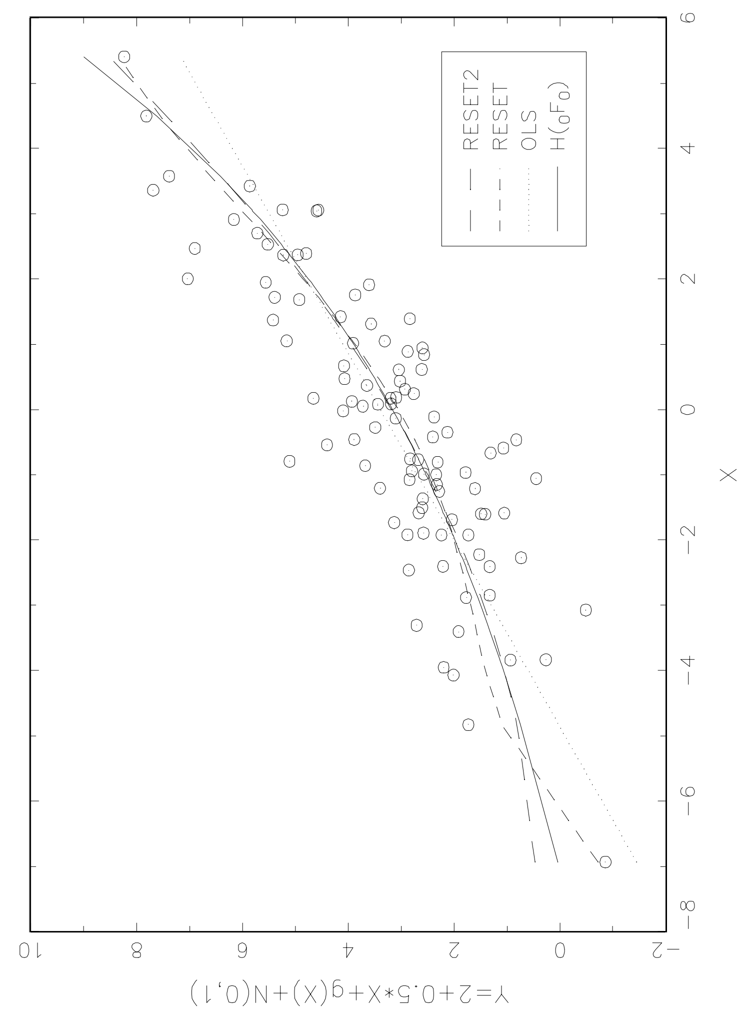

Figure 2: RESET, RESET2 and $\mathrm{H}({}_0F_0)$ fits; $N = 100$; $X_j \sim \mathrm{N}(0, 5)$; omission $g\,(X_j) = \exp\,(0.3X_j)$; circles represent generated points

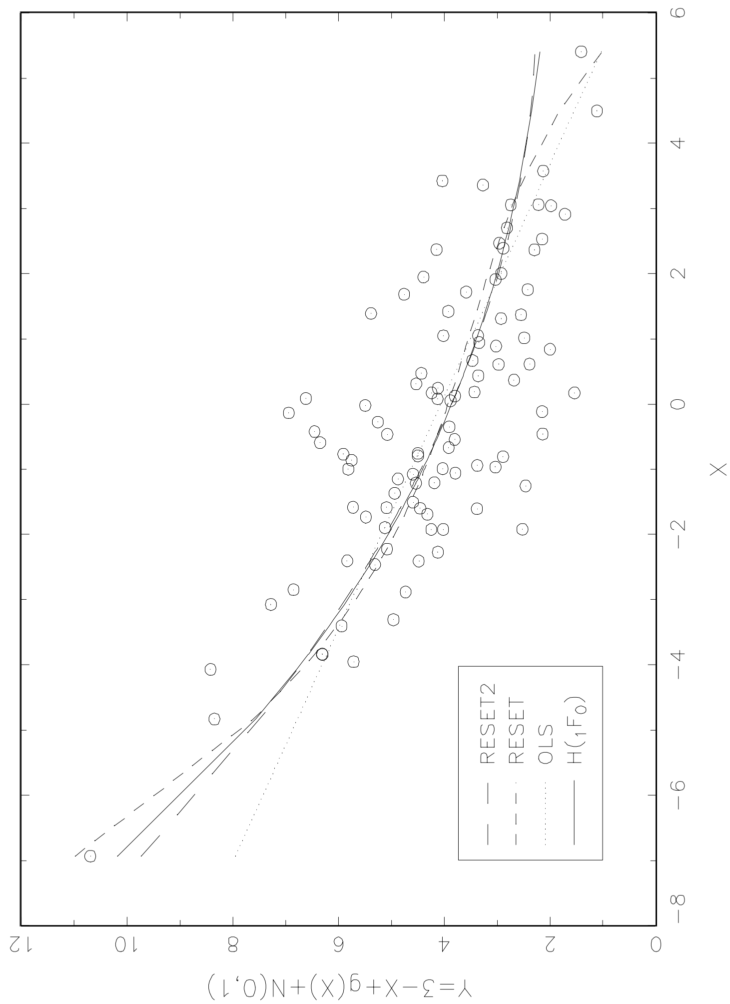

Figure 3: RESET, RESET2 and $H({}_1F_0)$ fits; $N = 100$; $X_j \sim N(0, 5)$; omission $g(X_j) = (1 + 0.2X_j)^2$; circles represent generated points

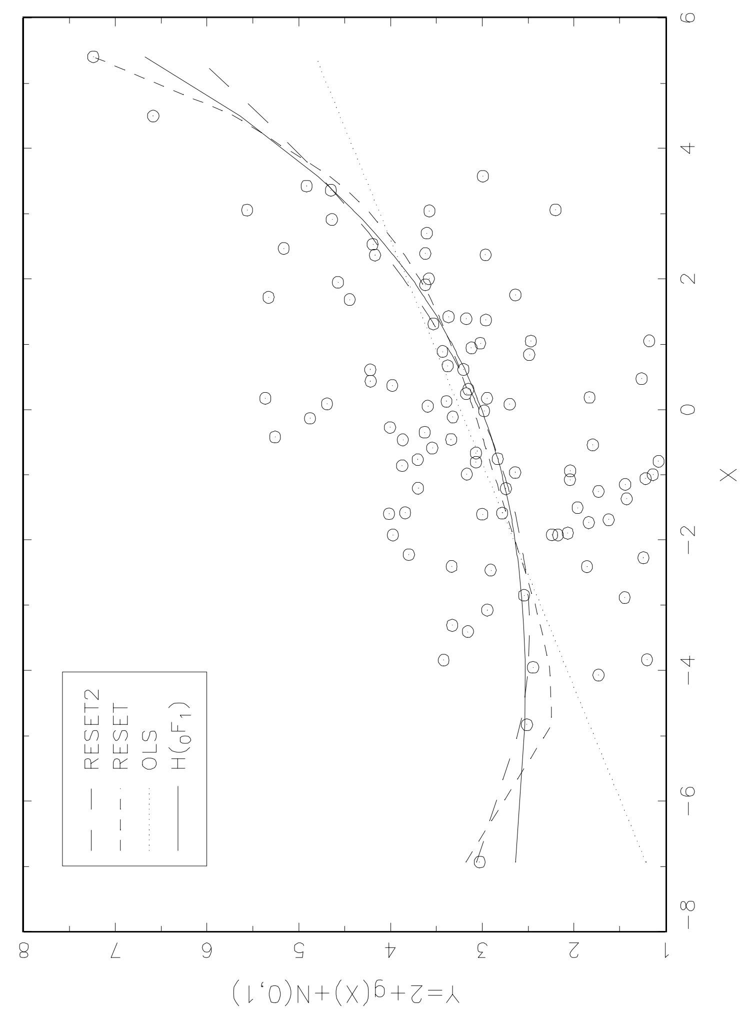

Figure 4: RESET, RESET2 and H($_0F_1$) fits; $N = 100$; $X_j \sim \mathrm{N}(0, 5)$; omission $g(X_j) = {}_0F_1(2; 0.7X_j)$; circles represent generated points

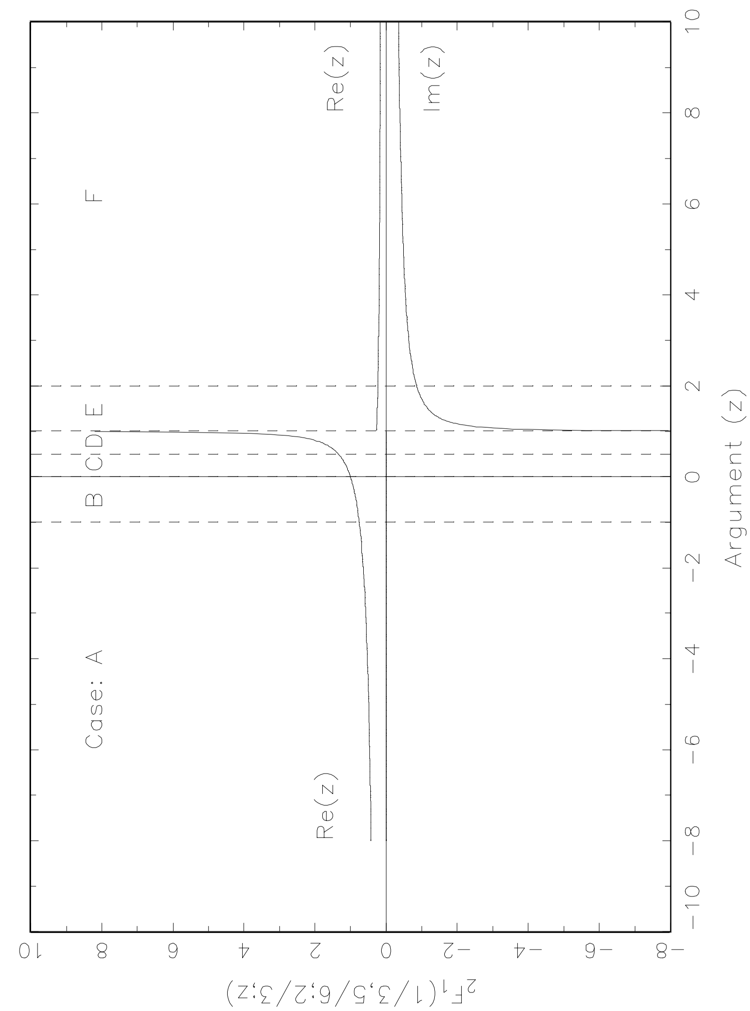

Figure 5: Generation of ${}_2F_1\left(\frac{1}{3}, \frac{5}{6}; \frac{2}{3}; z\right)$

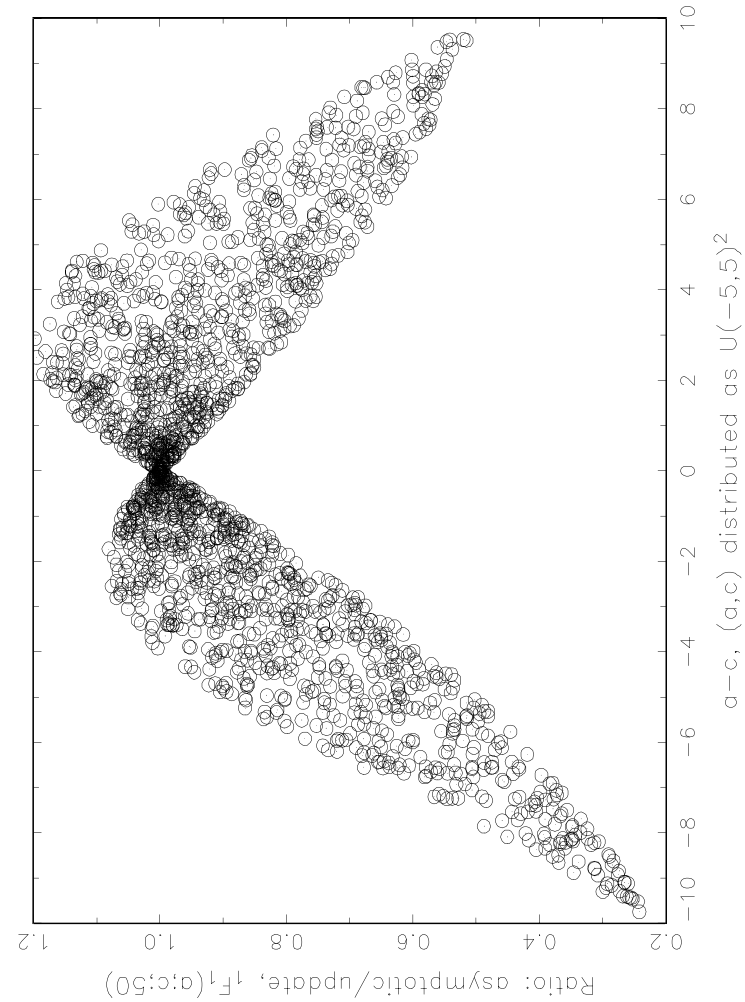

Figure 6: "Angel's wings", illustrates difficulty of using asymptotics for the $_1F_1$; $2,000$ points displayed

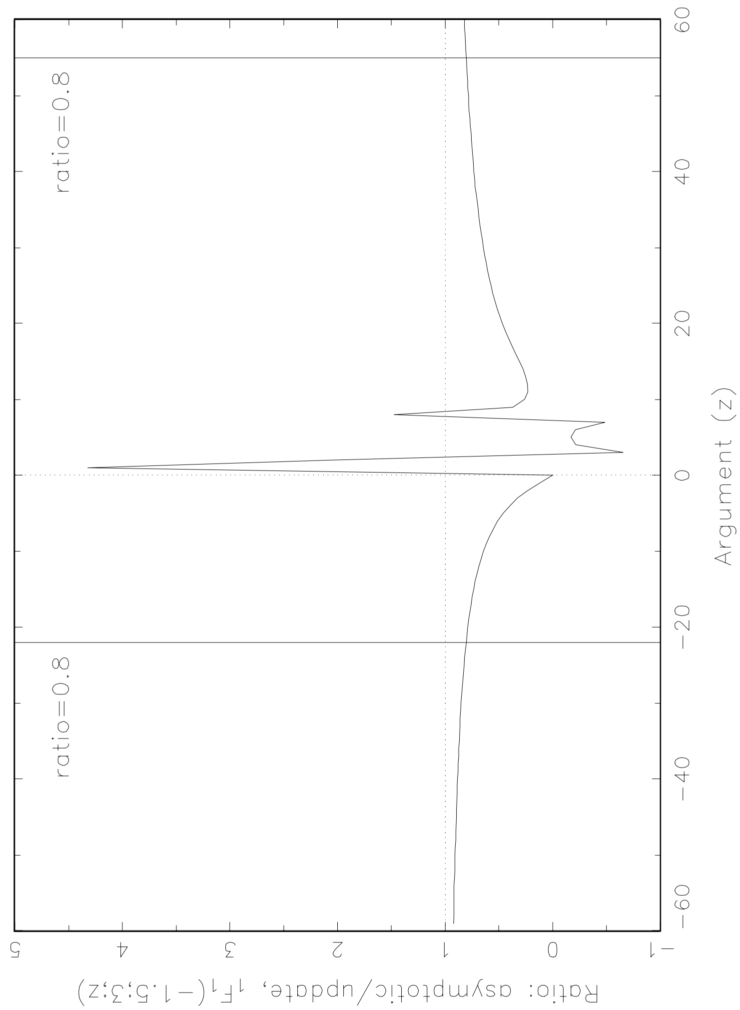

Figure 7: Generation of ${}_1F_1\,(-1.5;3;z)$ using Kummer Transform for $z<0$; ratio of asymptotic expression to linear update result; vertical bold lines indicate arguments for which ratio ≈ 0.8.