Reduced Forms and Weak Instrumentation*

Peter C. B. Phillips
Yale University, University of Auckland,
Singapore Management University & University of Southampton

December 9, 2015

Abstract

This paper develops exact finite sample and asymptotic distributions for a class of reduced form estimators and predictors, allowing for the presence of unidentified or weakly identified structural equations. Weak instrument asymptotic theory is developed directly from finite sample results, unifying earlier findings and showing the usefulness of structural information in making predictions from reduced form systems in applications. Asymptotic results are reported for predictions from models with many weak instruments Of particular interest is the finding that, in unidentified and weakly identified structural models, partially restricted reduced form predictors have considerably smaller forecast mean square errors than unrestricted reduced forms. These results are related to the use of shrinkage methods in system-wide reduced form estimation.

Keywords: Endogeneity, Exact distribution, Finite sample theory, Moment existence, Partial identification, Reduced form, Structural equation, Unidentified structure, Weak instrument.

JEL classifications: C23, C32

1 Introduction

In pioneering work on reduced form estimation, Maasoumi (1978) showed how to construct a modified Stein-like estimator of the reduced form coefficients in a linear simultaneous equations system. This technique cleverly combined information from restricted and unrestricted estimation via the medium of the system-wide overidentification test statistic for the structural restrictions. Innovative in its use of a shrinkage method driven by structural information in a reduced form context, this procedure had a further motivating feature that

*A preliminary version of this paper was presented at the Emory University Conference honoring Esfandiar Maasoumi on 15 November 2014. The author thanks Aman Ullah and three referees for helpful comments and acknowledges support of the NSF (USA) under Grant SES 12-58258 and Grant NRF-2014S1A2A2027803 from the Korean Government.

it ensured a thinning of the tails of the finite sample distribution of the reduced form coefficients, thereby enhancing their utility in applications such as forecasting. The latter advantage was seen at the time to break new ground in reduced form estimation because of the discovery, only a few years earlier, that many commonly used structural equation estimators such as two-stage and three-stage least squares methods produced reduced form estimators that had heavy tails and possessed no finite integer moments.

The second advance in Maasoumi's paper was the customized nature of the construction of the reduced form estimator. This estimator used Stein-like shrinkage of the restricted reduced form estimator toward an unrestricted reduced form estimator with a shrinkage tuning factor based on the outcome of a statistical test of the restrictions. The methodology followed Sargan's (1958, 1959) early work on overidentification testing and the test statistic mirrored Malinvaud's (1966) system-wide test of structural equation restrictions in a simultaneous equations setting.

The present work has a similar focus on reduced form estimation. We look particularly at cases where there is weak structural identification and explore the implications of such weakness for reduced form coefficient estimation and prediction. The methodology we pursue follows the approach developed originally in Phillips (1989) of relating exact finite sample distribution theory to asymptotic theory in conditions of underidentification. We provide an overview of this approach in the first part of the paper and then proceed to develop a similar analysis for reduced form coefficient estimation.

The plan of the paper is as follows. Section 2 reviews some exact finite sample theory for instrumental variable (IV) structural estimation and illustrates its use in deriving large sample asymptotics, asymptotic expansions, limit theory under weak identification, and many instrument asymptotics. While the capabilities of this approach are well understood in structural model estimation, the mechanism has been seldom used in other applications. Section 3 uses the approach to develop exact finite sample theory and asymptotics for reduced form estimators and predictors that take advantage of structural information even in the context of poorly identified structural systems. Cases of weak instrumentation and the use of many weak instruments are studied. Simulation exercises are reported to highlight the practical implications of the findings. The conclusion in Section 4 offers some more general reflections on exact theory and reduced forms. Proofs are given in the Appendix.

2 Exact Distribution Theory and its Asymptotic Implications

2.1 A Prototypical Simultaneous System

We write the linear simultaneous equations system in the following form

$$By_t + Cz_t = Ax_t = u_t, \quad t = 1, ..., n \tag{1}$$

where $A = [B, C]$ is an $m \times (m + k)$ matrix of unknown structural coefficients, $x_t = (y_t', z_t')'$ is a vector of m endogenous variables (y_t) and k exogenous variables (z_t), and $u_t \sim_{iid} N(0, \Omega_u)$ is a vector of serially independent disturbances on the structural equations. Gaussianity is assumed for the development of a finite sample theory but is not needed for the asymptotics (c.f. Phillips, 1989). The reduced form of system (1) is written as

$$y_t = P z_t + v_t, \text{ with } P = -B^{-1}C \text{ and } v_t \sim_{iid} N\left(0, \Omega_v = B^{-1}\Omega_u B'^{-1}\right). \quad (2)$$

The notation in (1) and (2) mirrors that used in Maasoumi (1978) which follows the Sargan LSE lectures (Sargan, 1988a) and the Malinvaud (1966) textbook tradition.

The system (1) is assumed to have sufficient restrictions to identify, at least apparently[1], the structural coefficients A. These restrictions may take the form of analytic constraints on the coefficients A or direct functional representation of the matrix A in terms of a subset θ of parameters such as $A = A(\theta)$. Identification of a particular structure, as distinct from apparent identification, requires additional conditions, of course, but these are not necessary for estimation (Phillips, 1989). When interest centers on the reduced form, the matrix P of relevant coefficients in (2) may be estimated directly by unrestricted reduced form least squares giving $\hat{P}$ or indirectly by $P^\dagger = -B^{\dagger-1}C^{\dagger-1}$ using the structural coefficient estimates $\left[B^\dagger, C^\dagger\right]$ obtained by a method such as three stage least squares (3SLS).

(a) Maasoumi Reduced Form Estimation

We follow Maasoumi (1978) and assume the restricted estimate $P^\dagger$ is obtained by 3SLS. Throughout the paper we denote observation matrices by capitals, for instance $Z = [z_1, ..., z_n]'$, and projection matrices by $P_Z = Z(Z'Z)^{-1}Z'$ and $Q_Z = I_n - P_Z$, so that the unrestricted reduced form estimator is $\hat{P} = Y'Z(Z'Z)^{-1}$. Maasoumi suggested a customized reduced form estimator of P that combined $P^\dagger$ with $\hat{P}$ using weights delivered by a shrinkage tuning factor determined by the outcome of a statistical test of the restrictions involved in the full specification of (1).

The test proposed by Maasoumi was based on a Wald statistic, given in tensor trace form as

$$\phi^\dagger = tr\left[\hat{\Omega}_v \left(\hat{P} - P^\dagger\right) Z'Z \left(\hat{P} - P^\dagger\right)'\right], \quad (3)$$

where $\hat{\Omega}_v = n^{-1}Y'Q_Z Y$ is the usual unrestricted residual moment matrix estimate of Ω_v. The statistic $\phi^\dagger$ follows Malinvaud (1966, chapter 9.5) and can equivalently be written as an overidentification test statistic analogous to that

[1] By apparent identification, we mean that order conditions enumerating the number and form of the restrictions appear, prima facie, to indicate identification, but without the assurance of supporting rank conditions that confirm relevance, to use the terminology of Phillips (1989).

developed originally in the work of Sargan (1958, 1959). Notably, under correct specification we have the asymptotic distribution $\phi^\dagger \sim_a \chi^2_N$, where N is the total number of overidentifying degrees in the structural system (1). In large samples, the restrictions are rejected if $\phi^\dagger$ exceeds the test critical value, although in small samples the test typically overrejects, as was already well known in the 1970s. Maasoumi used this feature of the test to construct a combined estimator of the reduced form matrix

$$P^* = \lambda P^\dagger + (1-\lambda)\hat{P} = P^\dagger + (1-\lambda)\left(\hat{P} - P^\dagger\right). \tag{4}$$

If C_p is a chosen critical value for the test, then the weight $\lambda \in [0,1]$ is chosen so that

$$\lambda = \begin{cases} 1 & \text{if } \phi^\dagger \leq C_p \text{ (hypothesis is accepted)} \\ \frac{\phi_2}{\phi^\dagger} & \text{if } \phi^\dagger > C_p \text{ (hypothesis is rejected)} \end{cases}$$

for some $\phi_2 \leq C_p$. With this rule, the combined estimator has the form

$$P^* = P^\dagger + \mathbf{1}_{\{C_p,\infty\}}\left(1 - \frac{\phi_2}{\phi^\dagger}\right)\left(\hat{P} - P^\dagger\right),$$

with the indicator $\mathbf{1}_{\{C_p,\infty\}} = 1$ if $\phi^\dagger > C_p$ achieving a switch from the restricted estimate $P^\dagger$ to the combined estimate P^* that shrinks $P^\dagger$ toward the unrestricted $\hat{P}$ when the test rejects the restrictions. Maasoumi shows that the combined estimator P^* has finite integer order moments to order $n - m - k$. If $C_p \to \infty$ as $n \to \infty$ at an appropriate rate then there are no false positives in the specification test and P^* has the same asymptotic distribution as $P^\dagger$, thereby capturing any implied advantages from the structural information for forecasting and other uses of the reduced form.

This innovative approach to reduced form estimation has several advantages. First, the combined estimator P^* has finite moments and thin tails compared with the 3SLS estimate $P^\dagger$, which typically has no finite sample moments and a heavy tailed distribution (Sargan, 1976/1988). Second, under broad conditions on the construction of the test, P^* has the same limit theory as $P^\dagger$, so it carries all the advantages of the additional structural information, when data deems this information correct. Third, P^* is a combined OLS-3SLS estimator, has the implicit advantages of an averaging estimator, and is more readily computed than the alternative full information maximum likelihood estimator, a significant gain at the time, particularly for prediction purposes where regular updating is required.

Reduced form estimation using this technique offered good prospects, as were quickly recognized by Edmond Malinvaud in his 1980 public address to the Econometric Society World Congress in Aix en Provence. Nonetheless, the impact of this methodology on empirical research and, in particular, on forecasting has been slow in arriving, although methods that do not use structural information have become common. For instance, shrinkage methods were suggested during the 1980s for use in Bayesian vector autoregressions to achieve parsimony in forecasting (Doan et al, 1984) and data-driven approaches to shrinkage factor selection have been developed (Phillips, 1996). In the last decade, generalized approaches to shrinkage estimation such as Lasso (Tibshirani, 1996), adaptive Lasso (Zou, 2006), bagging (Breiman, 1996) and their many variants, have begun to impact empirical work in economics. Some of these methods now use potential structural information or long-run behavior in adapting the Lasso mechanism (Liao and Phillips, 2015). Recent work on averaging estimators has also increased the popularity of these techniques (among others, see Fan and Ullah, 1999; Hansen, 2007, 2009, 2014; Iglesias and Phillips, 2012; Kotlyarova and Zinde-Walsh, 2006).

(b) Moment Exponents and Heavy Tails

By virtue of its construction, the restricted estimator $P^{\dagger} = B^{\dagger -1} C^{\dagger} = \frac{adj\left(B^{\dagger}\right)}{\det\left(B^{\dagger}\right)} C^{\dagger}$ is a rational function of the elements of the structural coefficient estimates $\left[B^{\dagger}, C^{\dagger}\right]$. Sargan (1976/1988) proved that such reduced form estimates (obtained by least squares and related structural equation methods like 3SLS) typically have maximal moment exponent of unity, so that no finite integral order moments exist. The distributions of such estimates, and hence those of the induced forecasts, inevitably therefore have heavy tails.

The elegant argument used by Sargan relies on the ratio form of $P^{\dagger}$ and the fact that most structural equation estimates $\left[B^{\dagger}, C^{\dagger}\right]$ have a probability density that is positive as $\det(B)$ passes through the origin. Moments of such reduced form estimates may therefore be bounded below as follows

$$\mathbb{E}\left(\left\|P^{\dagger}\right\|^{r}\right) \geq \int_{N_{(|\det(B)| \leq \delta)}} \frac{\|adj\,(B)\|^{r}}{|\det(B)|^{r}} \|C\|^{r} f_{B^{\dagger}, C^{\dagger}}(B, C)\, dB \times dC, \tag{5}$$

where $f_{B^{\dagger}, C^{\dagger}}(B, C)$ is the density of the unrestricted elements of (B, C) in (1) with respect to Lebesgue measure $dB \times dC$ on the space of those elements, $N_{(|\det(B)| \leq \delta)} = \{(B, C) : |\det(B)| \leq \delta\}$ for some small $\delta > 0$, and $\|\cdot\|$ is a matrix norm. When $f_{B^{\dagger}, C^{\dagger}}(B, C) > 0$ on $N_{(|\det(B)| \leq \delta)}$, then the integral in (5) does not converge and $\mathbb{E}\left(\left\|P^{\dagger}\right\|^{r}\right) = \infty$ for all $r \geq 1$. So the maximal moment exponent is unity and no finite integral moments of the elements of $P^{\dagger}$ exist. Exceptions can occur for triangular systems where $\det(B)$ is constant and non zero (provided the elements of $\left(B^{\dagger}, C^{\dagger}\right)$ also have moments to high enough order) and in some other special situations.

We illustrate with the simple example of a single structural equation where $y_t' = (y_{1t}, y_{2t})$ with one exogenous variable z_t combined with an identity (see Phillips, 2009), so that

$$B = \begin{bmatrix} 1 & -b \\ -1 & 1 \end{bmatrix}, \; C = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \; \Omega_u = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix},$$

and $\det(B) = 1 - b$. Suppose $b^{\dagger}$ is the ordinary least squares estimate of b obtained from the regression of y_{1t} on y_{2t}. In this case, the density of $b^{\dagger}$ is known to be supported and non zero on the whole real axis, so that the density is positive in the neighborhood $b^\dagger \sim 1$. It follows that the corresponding estimates of the reduced form coefficients

$$P^\dagger = \frac{1}{1-b^\dagger}\begin{bmatrix} 1 & b^\dagger \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

will have no integer finite sample moments and their distributions will have heavy tails. By contrast, in this example, the least squares reduced form coefficient estimate $\hat{P}$ is normally distributed and has Gaussian tails. Moreover, if the structural equation is estimated by maximum likelihood (which happens to correspond to indirect least squares in this particular case) giving $\tilde{b}$, then it is known that the finite sample density of estimate of $\tilde{b}$ has zero density at $b = 1$ (see Phillips, 2009)[2]. The corresponding reduced form estimates are just the least squares estimates and have finite integer moments and Gaussian tails.

This example illustrates Sargan's (1973) important finding that FIML estimates of the reduced form coefficients do have finite integer order moments. The reason maximum likelihood avoids the problem of a positive density of $\det\left(\tilde{B}\right)$ at points where $\det B = 0$ is that system maximum likelihood recognizes the simultaneous equations nature of the system. So existence of the reduced form (that is, $\det(B) \neq 0$) is critical to the formation of the likelihood, which ensures that zero density is attached to $\det\left(\tilde{B}\right) = 0$, thereby avoiding the problem of heavy tails in the reduced form coefficients.

2.2 Structural Equation Estimation and Weak IV

We now consider a single structural equation of (1) containing n observations on $m+1$ endogenous variables $[y_1, Y_2]$ written in conventional observation form as

$$y_1 = Y_2\beta + Z_1\gamma + u, \tag{6}$$

with conformably partitioned reduced form (where the integer symbols above the matrices indicate column dimension)

$$Y = [\overset{1}{y_1}, \overset{m}{Y_2}] = [\overset{K_1}{Z_1}, \overset{K_2}{Z_2}]\begin{bmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22} \end{bmatrix} + [v_1, V_2] = Z\Pi + V, \tag{7}$$

and corresponding identifiability relations

[2] The density of $\tilde{b}$, first derived in Bergstrom (1962) and used later in Nelson and Startz (1990), has the form

$$f_{\tilde{b}}(b) = \sqrt{\frac{\lambda_n}{2\pi\sigma^2}}\frac{1-\beta}{(1-b)^2}\exp\left\{-\frac{n}{2\sigma^2}\left(\frac{b-\beta}{1-b}\right)^2\right\}$$

where $\lambda_n = \sum_{t=1}^n z_t^2$ is the noncentrality parameter in this system. Clearly $\lim_{b\to 1} f_{\tilde{b}}(b) = 0$. By contrast, the density of the least squares estimator of b is positive at $b = 1$, as shown in Phillips (2009). The density $f_{\tilde{b}}(b)$ is also well known to be bimodal (Phillips and Wickens, 1978; Phillips and Hajivassilou, 1984, Nelson and Startz, 1990; Phillips, 2006, Fiorio et al., 2010). Bimodality is particularly evident in the presence of strong endogeneity (as in this example) when the instruments are weak (Phillips, 2006).

$$\pi_{11} - \Pi_{12}\beta = \gamma; \quad \pi_{21} - \Pi_{22}\beta = 0. \tag{8}$$

Our focus of attention in what follows will be on instrumental variable (IV) estimation of the structural equation (6) in conjunction with the estimation of the reduced form. With no loss of generality, it is convenient to assume standardizing transformations (see Phillips, 1983) are applied so that $n^{-1}Z'Z = I_K$ where $K = K_1+K_2$, and the covariance matrix of the elements of $V = [v_1, V_2]$ is $I_{n(m+1)}$. Since $u = v_1 - V_2\beta$, it follows that $\mathbb{E}(uu') = \left(1+\beta'\beta\right) I_n$ and the structural equation error variance is $1+\beta'\beta$.

We start with the estimation of (6) using an IV observation matrix H selected from the included and excluded (that is, excluded from the structural equation (6)) exogenous variables $Z = [Z_1, Z_2]$ according to the scheme

$$H = [Z_3, Z_1] = [\overset{m+L}{Z_2 S}, \overset{K_1}{Z_1}].$$

Here S is a selector matrix that selects $m+L \leq K_2$ instruments from the excluded covariates Z_2 and the integer $L \geq 0$ is the degree of (apparent) overidentification or surplus instrumentation. The exogenous variable matrix Z is taken as fixed and of full rank K, although other options are possible and employed later in the paper. The IV estimator of β can be written as

$$\beta_{IV} = \left[Y_2'\left(P_H - P_{Z_1}\right)Y_2\right]^{-1} Y_2'\left(P_H - P_{Z_1}\right)y_1 = \left[Y_2'CC'Y_2\right]^{-1} Y_2'CC'y_1, \tag{9}$$

where $C = Q_{Z_1}Z_3\left(Z_3'Q_{Z_1}Z_3\right)^{-1/2}$ is an $n \times (m+L)$ matrix of orthonormal vectors, so that $C \in V_{m+L,n}$, the Stiefel manifold of $m+L$ orthonormal vectors of dimension n.

We now briefly review the exact distribution theory of β_{IV} and associated asymptotic expansions that facilitate some of the remaining developments. This distribution theory reveals in a simple way the weak IV asymptotics that apply when $n \to \infty$ and the relations (8) are only weakly identifying in the sense that $\Pi_{22} = n^{-1/2}\Pi_{22}^{*} = O\left(n^{-1/2}\right)$. In such cases, the identifiability relations (8) provide only limited information about β even as $n \to \infty$ and this uncertainty manifests itself in the asymptotics. The theory used here draws from Phillips (1980, 1989) and is used as a bridge to the reduced form theory derived later in the paper. Some results on many weak IV asymptotics are also provided.

2.3 IV Exact Distributions, Asymptotics and Expansions

Under Gaussianity and standardizing transforms, the error matrix V is matrix normal[3] $N_{n,m+1}\left(0, I_{n(m+1)}\right)$ and the data matrix

$$C'Y \sim_d N_{m+L,m+1}\left(M', I_{(m+L)n}\right), \tag{10}$$

[3] We use the notation $V \sim_d N_{n,m+1}\left(0, I_{n(m+1)}\right)$ to signify that the matrix V is normally distributed, i.e., the $n(m+1)$ vector $vec(V) \sim_d N\left(0, I_{n(m+1)}\right)$.

with mean vector

$$\begin{aligned} M' &= \mathbb{E}(C'Y) = (Z_3'Q_{Z_1}Z_3)^{-1/2} Z_3'Q_{Z_1}Z\Pi = n^{1/2}(S'S)^{-1/2} S'[\pi_{21}, \Pi_{22}] \\ &= n^{1/2}(S'S)^{-1/2} S'\Pi_{22}[\beta, I_m] = n^{1/2}\bar{\Pi}_{22}[\beta, I_m], \end{aligned}$$

where $\bar{\Pi}_{22} = (S'S)^{-1/2} S'\Pi_{22}$ is $(m+L) \times m$ with rank at most m. The sample moment matrix $Y'CC'Y$ is then distributed as noncentral Wishart of dimension $m+1$ with covariance matrix I_{m+1}, noncentrality matrix

$$MM' = n \begin{bmatrix} \beta' \\ I_m \end{bmatrix} \bar{\Pi}_{22}'\bar{\Pi}_{22}[\beta, I_m], \tag{11}$$

whose rank is at most m, and degrees of freedom $m+L$. This distribution is written as $W_{m+1}(m+L, I_{m+1}, MM')$ and, conformable with the partition of the structural equation (9), we may write the matrix quadratic form

$$Y'CC'Y = \begin{bmatrix} \overset{1}{a_{11}} & \overset{m}{a_{12}} \\ a_{21} & A_{22} \end{bmatrix} \sim_d W_{m+1}(m+L, I_{m+1}, MM'). \tag{12}$$

In terms of these components, the IV estimator (9) is $\beta_{IV} = A_{22}^{-1}a_{21}$, a matrix quotient of the components of the Wishart matrix (12). The matrix $n\bar{\Pi}_{22}'\bar{\Pi}_{22}$ in (11) is called the concentration parameter matrix and is instrumental in determining the (finite sample and asymptotic) properties of β_{IV} because information about β is transmitted to the density (12) and hence to the density of β_{IV} through the noncentrality matrix MM' by virtue of the matrix $n\bar{\Pi}_{22}'\bar{\Pi}_{22}$ and its behavior as $n \to \infty$.

The exact distribution of β_{IV} was obtained in Phillips (1980) and the density has the following series form

$$\begin{aligned} f_{IV}(r) &= \frac{\operatorname{etr}\left\{-\frac{n}{2}\left(I + \beta\beta'\right)\bar{\Pi}_{22}'\bar{\Pi}_{22}\right\}\Gamma_m\left(\frac{L+m+1}{2}\right)}{\pi^{m/2}\left[\det\left(I_m + rr'\right)\right]^{(L+m+1)/2}} \\ &\times \sum_{j=0}^{\infty} \frac{\left(\frac{L}{2}\right)_j \left[\left(\frac{n}{2}\beta'\bar{\Pi}_{22}'\partial_W^a\bar{\Pi}_{22}\beta\right)^j \det\left(I_{m+L} + W\right)^{\frac{L-1}{2}+j}}{j!\Gamma_n\left(\frac{L+m}{2}+j\right)} \\ &\times {}_1F_1\left(\begin{array}{c} \frac{L+m+1}{2}, \frac{L+m}{2}+j; \\ \frac{n}{2}\left(I_{m+L}+W\right)\bar{\Pi}_{22}\left(I_m + \beta r'\right)\left(I_m + rr'\right)^{-1}\left(I_m + r\beta'\right)\bar{\Pi}_{22}' \end{array}\right)\Bigg]_{W=0} \end{aligned} \tag{13}$$

where $\operatorname{etr}\{\cdot\}$ signifies $\exp\{\operatorname{trace}(\cdot)\}$, $\partial_W^a = \operatorname{adj}(\partial/\partial W)$ is the polynomial differential operator obtained by taking the adjoint matrix of the matrix operator $\partial/\partial W$, ${}_1F_1(a, b; U)$ denotes a confluent hypergeometric function with matrix argument U and parameters (a, b), Γ is the gamma function, Γ_m denotes the multivariate gamma function for $m > 1$, and the notation $(a)_j = a(a+1)\ldots(a+j-1) = \Gamma(a+j)/\Gamma(a)$ is the Pochhammer forward factorial function. Readers are referred to Phillips (1980, 1983) and Muirhead (1982) for background information on the matrix spaces involved, the multivariate methods employed, and further details of the special functions ${}_1F_1$ and Γ_m.

The exact density (13) was obtained under Gaussian errors but may be extended in some cases, for example by using Gram Charlier representations of more general error distributions. However, the results implied by (13) turn out to be important and relevant in many other cases including those of non-Gaussian errors, as detailed below. Importantly, the expression $f_{IV}(r)$ for the density facilitates many further developments and specializations, including some helpful asymptotic results. We mention particularly the following.

(i) Asymptotic theory as $n \to \infty$ follows immediately by expanding the matrix argument ${}_1F_1$ function in (13) as $n \to \infty$.

(ii) Higher order asymptotics and Laplace approximations may also be obtained in the strong instrument case where $\bar{\Pi}_{22}$ has full rank. These are delivered simply by utilizing higher order expansions of the ${}_1F_1$ function as $n \to \infty$. In particular, the Laplace approximation to the density (13) is the much simpler expression

$$f_{IV}(r) = \frac{n^{m/2}\text{etr}\left\{-\frac{n}{2}\frac{\bar{\Pi}_{22}'\bar{\Pi}_{22}(r-\beta)(r-\beta)'}{1+r'r}\right\}\left(\det\left(\bar{\Pi}_{22}'\bar{\Pi}_{22}\right)\right)^{1/2}\left(1+\beta' r\right)^{L+1}}{(2\pi)^{m/2}\left(1+r'r\right)^{(L+m+2)/2}\left(1+2\beta' r-\beta'\beta\right)^{L/2}} \times \left\{1+O\left(n^{-1}\right)\right\}, \tag{14}$$

which delivers immediately the density of the limit distribution of $X_n = \sqrt{n}\left(\beta_{IV}-\beta\right)$ as

$$f_X(x) = \frac{e^{-\frac{1}{2}\frac{x'\bar{\Pi}_{22}'\bar{\Pi}_{22}x}{1+\beta'\beta}}\left(\det\left(\bar{\Pi}_{22}'\bar{\Pi}_{22}\right)\right)^{1/2}}{(2\pi)^{m/2}\left(1+\beta'\beta\right)^{m/2}} \equiv N\left(0,\left(1+\beta'\beta\right)\left(\bar{\Pi}_{22}'\bar{\Pi}_{22}\right)^{-1}\right). \tag{15}$$

(iii) Asymptotic theory in the unidentified case where $\bar{\Pi}_{22}=0$ is also easily obtained without the Gaussianity assumption on the error matrix V. We need only take the leading term of the series representation of the exact density $f_{IV}(r)$, which gives the following (scaled) multivariate t distribution[4] with $L+1$ degrees of freedom and density

$$\frac{\Gamma_m\left(\frac{L+m+1}{2}\right)}{\pi^{m/2}\Gamma_n\left(\frac{L+m}{2}\right)\left[\det\left(I_m+rr'\right)\right]^{(L+m+1)/2}} = \frac{\Gamma\left(\frac{L+m+1}{2}\right)}{\pi^{m/2}\Gamma\left(\frac{L+1}{2}\right)\left[1+r'r\right]^{(L+m+1)/2}}, \tag{16}$$

and note that a central limit theorem (CLT) produces the required Gaussianity of (10) without assuming Gaussianity for V. Observe further that when $L=0$, the limiting density is multivariate Cauchy and has no integer finite sample moments. Thus, IV estimation under conditions of apparent

[4] Strictly speaking the density (16) is proportional to a multivariate t distribution. In particular, the distribution given by the density (16) is the distribution of $t_q/q^{1/2}$ with $q = L+1$ where t_q is multivariate t with q degrees of freedom. See Phillips (1989, theorem 2.1).

just identification has a heavy tailed density. Asymptotic theory for statistical tests, including general Wald test of restrictions on the structural coefficients β, are also easily deduced.

(iv) Asymptotics in the weakly identified case also follow immediately for reduced form matrices that are local to zero such as $\bar{\Pi}_{22} = n^{-1/2}\bar{\Pi}_{22}^* = O\left(n^{-1/2}\right)$, for some matrix $\bar{\Pi}_{22}^*$ of localizing coefficients. In this case, the limit distribution of β_{IV} as $n \to \infty$ is just the exact density $f_{IV}(r)$ of β_{IV} given above in (13) with the simple replacement of $\bar{\Pi}_{22}$ by $n^{-1/2}\bar{\Pi}_{22}^*$. Similarly, asymptotic theory for statistical tests concerning β follows, just as in (iii). In all these cases, we simply can use the martingale CLT in Philllips (1989) to deliver the limit theory without assuming Gaussianity.

(v) The case of many weak instruments may also be explored using the exact theory. In particular, suppose the number of surplus instruments (or degree of overidentification) $L \to \infty$. Then, under certain regularity conditions that include the expansion rate of L, β_{IV} is consistent with a slower rate of convergence than $\sqrt{n}$ and the rate of convergence depends on the extent of the weakness in the instruments. But when $L \to \infty$ too rapidly relative to the expansion rate of the concentration matrix $\bar{\Pi}_{22}^{*\prime}\bar{\Pi}_{22}^*$, then β_{IV} is inconsistent.

Items (i)-(ii) are shown in Phillips (1980) and the various items in (iii) were proved in Phillips (1989), including the asymptotic theory of the associated statistical tests. We briefly demonstrate (iv) here as the methods are of some pedagogical interest in view of their simplicity. First observe that the replacement $\bar{\Pi}_{22} = n^{-1/2}\bar{\Pi}_{22}^*$ ensures that (10) holds with $M' = \bar{\Pi}_{22}^*[\beta, I_m]$ as $n \to \infty$ by the martingale CLT. In particular

$$C'Y - \mathbb{E}(C'Y) = C'V \sim_a N_{m+L,m+1}\left(0, I_{(m+L)(m+1)}\right), \tag{17}$$

because $C'V$ satisfies the stability and Lindeberg conditions, as shown in Phillips (1989, lemma 2.3), when the rows of V are stationary and ergodic martingale differences with conditional covariance matrix I_{m+1}. Next note that β_{IV} is a continuous function of $C'Y$ by virtue of (9) and the fact that the limiting matrix normal distribution (and the implied Wishart distribution of $Y_2'CC'Y_2$) is of full rank, confirming continuity of the mapping. The exact density $f_{IV}(r)$ derived under Gaussianity is then the limiting density of β_{IV} as $n \to \infty$ when the reduced form coefficients satisfy $\Pi_{22} = n^{-1/2}\Pi_{22}^*$ and the data are not necessarily Gaussian. That is, when the instrument matrix H is weak for the endogenous variables Y_2 in the structural equation (6), the limit theory as $n \to \infty$ reproduces the exact distribution theory under Gaussianity with all its associated parameter dependencies under the simple replacement of Π_{22} by $n^{-1/2}\Pi_{22}^*$. Note that the dependence of the density (13) on the sample size n is removed by this replacement.

This approach brings together the exact distribution theory and the weak instrument limit theory by virtue of the simple action of the CLT (17) from

Phillips (1989) and continuous mapping from the matrix quotient form of the estimator β_{IV}. The argument is identical in this non-central (weak instrument) case to that given in Phillips (1989) for the unidentified case. Results for statistical tests follow directly. Staiger and Stock (1998) developed results in (iv) assuming high level conditions that require a CLT such as (17) and considered k-class estimators as well as tests of overidentification and exogeneity.

Item (v) was considered in Chao and Swanson (2005) and the results may be obtained in the present framework. For instance, we may assume that the noncentrality matrix $\bar{\Pi}_{22}'\bar{\Pi}_{22}$ satisfies $n\lambda_{\min}\left(\bar{\Pi}_{22}'\bar{\Pi}_{22}\right) = O\left(n^{2\alpha}\right) \to \infty$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of its matrix argument, and $n^{1-2\alpha}\bar{\Pi}_{22}'\bar{\Pi}_{22} \to \Psi > 0$ for some given $\alpha \in \left(0, \frac{1}{2}\right)$ and for some positive definite limit matrix Ψ. Then the density of the centered and scaled estimator $X_n = n^{\alpha}\left(\beta_{IV} - \beta\right)$ has the form

$$
\begin{aligned}
f_{X_n}(x) &= \frac{e^{-\frac{1}{2}\frac{x'\left(\frac{n}{n^{2\alpha}}\bar{\Pi}_{22}'\bar{\Pi}_{22}\right)x}{1+\beta'\beta}}\left(\det\left(\frac{n}{n^{2\alpha}}\bar{\Pi}_{22}'\bar{\Pi}_{22}\right)\right)^{1/2}}{(2\pi)^{m/2}\left(1+\beta'\beta\right)^{m/2}}\left\{1 + O\left(\frac{L}{n^{\alpha}}\right)\right\} \\
&\to N\left(0, \sigma_u^2\Psi^{-1}\right), \text{ with } \sigma_u^2 = 1 + \beta'\beta,
\end{aligned}
\tag{18}
$$

as $\frac{1}{L} + \frac{L}{n^{\alpha}} \to 0$. So, under certain regularity conditions as the number of weak instruments grow, β_{IV} is consistent at a reduced rate and has a limiting normal distribution with variance matrix $\sigma_u^2\Psi^{-1}$ that depends on the limit $n^{1-2\alpha}\bar{\Pi}_{22}'\bar{\Pi}_{22} \to \Psi$ and the scalar $\sigma_u^2 = 1 + \beta'\beta$, which is the variance of the structural equation error u_t after standardizing transformations on the reduced form have been performed, as indicated earlier.

Recall that $\bar{\Pi}_{22} = \left(S'S\right)^{-1/2}S'\Pi_{22}$ is $(m+L)\times m$, so that the number of rows of $\bar{\Pi}_{22}$ expands as $L \to \infty$. Thus, depending on the extent of the weakness of the instruments (measured by the magnitude of the elements of $\bar{\Pi}_{22}$), when there is an expanding instrument set the matrix quadratic form $\bar{\Pi}_{22}'\bar{\Pi}_{22}$ can accumulate information at some rate that is related to the expansion rate of L. This rate is given by the exponent α that appears in the excitation condition $n\lambda_{\min}\left(\bar{\Pi}_{22}'\bar{\Pi}_{22}\right) = O\left(n^{2\alpha}\right) \to \infty$ and the regularity condition required is that L not grow too fast relative to the excitation rate, so that $L/n^{\alpha} \to 0$. At the limits of its domain of definition when $\alpha = \frac{1}{2}$, we have the usual strong instrumentation case for which β_{IV} is consistent at a $\sqrt{n}$ rate; and when $\alpha = 0$, we have conventional weak instrumentation under which the estimator β_{IV} is inconsistent. Other cases may also be considered and are investigated in Chao and Swanson (2005). A further example is considered later in this paper.

3 Reduced Form Exact Distribution Theory and Associated Asymptotics

We start by considering the same triangular structural system as that used above, viz.,

$$y_1 = Y_2\beta + Z_1\gamma + u, \tag{19}$$

$$Y_2 = Z_1\Pi_{12} + Z_2\Pi_{22} + V_2, \tag{20}$$

in which the second block of equations is already in reduced form. The complete reduced form is

$$\begin{aligned} [y_1 \ , Y_2] &= [Z_1\pi_{11} + Z_2\pi_{21} + v_1 \ , Z_1\Pi_{12} + Z_2\Pi_{22} + V_2], \\ &= [Z_1(\Pi_{12}\beta + \gamma) + Z_2\Pi_{22}\beta + v_1 \ , Z_1\Pi_{12} + Z_2\Pi_{22} + V_2], \end{aligned} \tag{21}$$

where $v_1 = u + V_2\beta$. For the development of an exact theory we assume, as earlier, that $V \sim_d N\left(0, I_{n(m+1)}\right)$ after standardizing transformations. The Gaussian assumption is relaxed in the asymptotic theory discussed later.

The system (21) solves the structural form and leads to a reduced form with restricted coefficients that involve the structural parameters. However, the coefficient $\pi_{11} = \Pi_{12}\beta + \gamma$ is effectively unrestricted because it has dimension $K_1 \times 1$ and the K_1- vector γ is also unrestricted. On the other hand, $\pi_{21} = \Pi_{22}\beta$ implies that the reduced form coefficients π_{21} are restricted by the requirement that $\pi_{21} \in \mathcal{R}(\Pi_{22})$, the range space of Π_{22}, and that the specific linear combination of Π_{22} involves β. Note particularly that if $\Pi_{22} = 0$ then $\pi_{21} = 0$, which imports information about Π_{22} from the reduced form for Y_2 directly into the reduced form for y_1 as a consequence of the structural model.

The system (21) is known as a partially restricted reduced form (Kakwani and Court, 1972) because it incorporates restrictions on the reduced form equations from a single structural equation. These restrictions may be used in estimation and they lead to the partially restricted reduced form estimator (Knight, 1977)

$$\begin{bmatrix} \tilde{\pi}_{11} = \hat{\Pi}_{12}\beta_{IV} + \gamma_{IV} = \hat{\pi}_{11} \\ \tilde{\pi}_{21} = \hat{\Pi}_{22}\beta_{IV} \end{bmatrix}, \text{ where } \begin{bmatrix} \hat{\Pi}_{12} \\ \hat{\Pi}_{22} \end{bmatrix} = (Z'Z)^{-1} Z'Y_2.$$

The first component $\tilde{\pi}_{11} = \hat{\pi}_{11}$ is unrestricted and is Gaussian with all its moments finite. It is the second component $\tilde{\pi}_{21} = \hat{\Pi}_{22}\beta_{IV}$ which is of primary interest because it carries the effects of the structural information into reduced form estimation. Knight's (1977) work showed that the partially restricted reduced form estimator has finite moments of all orders.

Since $\tilde{\pi}_{11} = \hat{\pi}_{11}$ is an unrestricted estimator, we need not be concerned with this component. In what follows, therefore, we eliminate the included exogenous variables from (19) and consider the structural equation

$$y_1 = Y_2\beta + u \tag{22}$$

with no included exogenous variables, and the associated reduced form

$$[y_1, Y_2] = Z[\pi_1, \Pi_2] + [v_1, V_2] = Z\Pi + V, \tag{23}$$

in which $\pi_1 = \Pi_2\beta \in \mathcal{R}(\Pi_2)$. Without loss of generality, we continue to assume that standardizing transformations are applied to (23) so that $Z'Z = nI_K$ and $V \sim_d N\left(0, I_{n(m+1)}\right)$.

3.1 Reduced Form Predictors

As in Maasoumi (1978), we are particularly interested in the effects of reduced form estimation on prediction. We concentrate on obtaining one period ahead forecast distributions for the first structural equation using restricted and unrestricted estimation. The forecast object is therefore $y_{1,n+1} = z'_{n+1}\pi_1 + v_{1,n+1}$, where $y_{1,n+1}$ is the forecast period value of the endogenous variable in (22) and the reduced form error $v_{1,n+1} \sim_d N(0,1)$ is independent of V. In what follows, we will maintain the Gaussian assumption to develop the exact theory and relax this assumption in developing the asymptotics.

(a) Exact Theory and Asymptotics in the Unidentified Case

We start by exploring the leading case where the submatrix $\Pi_{22} = 0$ and the structural coefficient β is unidentified. The exact theory for the reduced form is simpler for this case, just as it is for structural estimation theory, but suffices to reveal interesting features of the restricted reduced form estimator. Note that in this case $\pi_1 = \Pi_2\beta = 0$.

The two predictors (based on the unrestricted estimate $\hat{\pi}_1$ and the restricted estimate $\hat{\Pi}_2\beta_{IV}$ of π_1) are given by

$$\hat{y}_{1,n+1} = z'_{n+1}\hat{\pi}_1 \text{ and } \tilde{y}_{1,n+1} = z'_{n+1}\tilde{\pi}_1 = z'_{n+1}\hat{\Pi}_2\beta_{IV}.$$

With no loss of generality it is convenient to maintain the standardizing transformations and add the normalizing condition $z'_{n+1}z_{n+1} = 1$ on the forecast period exogenous variables. Then the unrestricted forecast

$$\hat{y}_{1,n+1} \sim_d N\left(z'_{n+1}\pi_1, z'_{n+1}\left(Z'Z\right)^{-1}z'_{n+1}\right) = N\left(z'_{n+1}\pi_1, \frac{1}{n}\right) \tag{24}$$

is unbiased with variance $1/n$. The distribution of the restricted forecast $\tilde{y}_{n+1}$ is more complex and involves the product $\hat{\Pi}_2\beta_{IV}$ of the unrestricted reduced form estimate $\hat{\Pi}_2 = \left(Z'Z\right)^{-1}Z'Y_2$ and the structural form IV estimate β_{IV}.

The exact densities of the two predictors of $y_{1,n+1}$ are given in the following result.

Theorem *Under Gaussianity, the stated conditions for the model (22) - (23), and in the leading case where $\pi_1 = \Pi_2\beta = 0$, the exact densities of the* *normalized unrestricted and restricted reduced form predictors* $\hat{y}^P_{1,n+1} = z'_{n+1}\left(\sqrt{n}\hat{\pi}_1\right)$ and $\tilde{y}^P_{1,n+1} = z'_{n+1}\left(\sqrt{n}\hat{\Pi}_2\beta_{IV}\right)$ *are given by*

$$f_{\hat{y}^P_{1,n+1}}(y) = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}, \tag{25}$$

$$f_{\tilde{y}^P_{1,n+1}}(y) = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\sum_{j=0}^{\infty}\frac{\left(\frac{K-m}{2}\right)_j\left(\frac{1}{2}\right)_j}{j!\left(\frac{K}{2}\right)_j}\,{}_1F_1\left(-j,\frac{1}{2};\frac{y^2}{2}\right), \tag{26}$$

where ${}_1F_1\left(-j,\frac{1}{2};\frac{y^2}{2}\right) = \sum_{k=0}^{j}\frac{(-j)(-j+1)\ldots(-j+k-1)}{k!\left(\frac{1}{2}\right)_k}\left(\frac{y^2}{2}\right)^k$ *is a terminating confluent hypergeometric series.*

The proof of (26) is given in the Appendix and relies on a simple conditioning argument. The idea is that, for the model (22) - (23), the restricted reduced form estimate has the following decomposition

$$\begin{aligned}\sqrt{n}\hat{\Pi}_2\beta_{IV} &= \sqrt{n}\left(Z'Z\right)^{-1}Z'Y_2\times\left(Y_2'CC'Y_2\right)^{-1}Y_2'CC'y_1\\ &= C'V_2\left(V_2'CC'V_2\right)^{-1/2}\times\left(V_2'CC'V_2\right)^{-1/2}\left(V_2'CC'v_1\right),\end{aligned} \tag{27}$$

where $C' = \left(Z'Z\right)^{-1/2}Z'$ and the component variates $C'V_2\left(V_2'CC'V_2\right)^{-1/2}$ and $\left(V_2'CC'V_2\right)^{-1/2}\left(V_2'CC'v_1\right)$ are independent. The standardized estimate $\sqrt{n}\hat{\Pi}_2\beta_{IV}$ therefore has a mixed normal distribution[5], from which the density can be evaluated by direct integration over the space $V_{m,K}$ of the $K\times m$ matrix variate $\Upsilon = C'V_2\left(V_2'CC'V_2\right)^{-1/2}$.

As noted in the statement of the theorem, the component factor ${}_1F_1\left(-j,\frac{1}{2};\frac{y^2}{2}\right)$ is a terminating hypergeometric series. This series is of special interest because it arises in the computation of moments of a non-central normal distribution. In particular, as shown in the proof of the theorem, if $X\sim N\left(\mu,\sigma^2\right)$ the even moment formula (e.g., Winkelbauer, 2014)

$$\mathbb{E}\left(X^{2j}\right) = \sigma^{2j}2^j\frac{\Gamma\left(j+\frac{1}{2}\right)}{\sqrt{\pi}}\,{}_1F_1\left(-j,\frac{1}{2},-\frac{\mu^2}{2\sigma^2}\right),$$

yields

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(s+iy)^2}{2}}s^{2j}ds = 2^j\left(\frac{1}{2}\right)_j\,{}_1F_1\left(-j,\frac{1}{2},\frac{y^2}{2}\right),$$

upon setting $\mu = iy$ and $\sigma = 1$.

[5] A matrix variate X has a (variance matrix) mixed normal distribution if the density of X is a compound distribution of the form $p_X(x) = \int_{A>0}\varphi(x,A)\,p(A)\,(dA)$ where $\varphi(x,A)$ is the matrix normal density with covariance matrix A and the integral is taken over the matrix space $A>0$ with respect to the invariant measure (dA) on the cone of positive definite matrices weighted by the probability density $p(A)$ of A. See Muirhead (1982) for further details.

When $K=m$, the series (26) truncates at the first term giving $f_{\tilde{y}^P_{1,n+1}}(y) = e^{-\frac{y^2}{2}}/\sqrt{2\pi} = f_{\hat{y}^P_{1,n+1}}(y)$. In this case, the structural model is (apparently) just identified and there are no effective restrictions on the reduced form, so that $\hat{\pi}_1 = \hat{\Pi}_2 \beta_{IV}$ and the predictors are equivalent (i.e., $\tilde{y}^P_{1,n+1} = \hat{y}^P_{1,n+1}$). In this case there is no gain from using the structural form estimate β_{IV} because the two reduced form estimates are equivalent.

The following moment results enable a straightforward comparison between the two predictors in terms of variance and forecast mean square error.

Corollary *Under the same conditions as above, the normalized unrestricted and restricted reduced form predictors have variances*

$$\mathbb{E}\left\{\left(\hat{y}^P_{1,n+1}\right)^2\right\} = 1, \text{ and } \mathbb{E}\left\{\left(\tilde{y}^P_{1,n+1}\right)^2\right\} = \frac{m}{K}. \tag{28}$$

When $K=2$ and $m=1$ we have $\mathbb{E}\left\{\left(\tilde{y}^P_{1,n+1}\right)^2\right\} = \frac{1}{2}$, so the variance of the restricted reduced form predictor is one half the variance of the unrestricted predictor $\hat{y}^P_{1,n+1}$. When $m=1$ and $K \geq 2$ we have a variance reduction to $1/K$ of the variance of the unrestricted reduced form predictor $\hat{y}_{1,n}$. The distribution of the restricted reduced form predictor is therefore much less dispersed than that of the unrestricted estimate. This finding is intriguing in light of the fact that the restrictions are only apparent in the present case. Indeed, the restricted predictor $\tilde{y}^P_{n+1}$ depends on structural estimation of an unidentified parameter and the structural estimator β_{IV} itself is inconsistent. Yet the partially restricted reduced form estimator substantially reduces variance in the predictor.

How is it, then, that the restricted estimator $\tilde{\pi}_1 = \hat{\Pi}_2\beta_{IV}$, which depends on the inconsistent estimator β_{IV}, can be less dispersed than the unrestricted estimator $\hat{\pi}_1$? The explanation is that $\tilde{\pi}_1$ is the product of two estimators $\hat{\Pi}_2$ and β_{IV}, both of which are centred on the origin and the first of which is consistent for $\Pi_2 = 0$. The product distribution enhances concentration around the origin, thereby reducing variance. In effect, when we use the information in β_{IV} in reduced form estimation we employ both $\hat{\Pi}_2$, which is centred on zero (just like the unrestricted RF estimate $\hat{\pi}_{21}$) and β_{IV}, which is also centered on zero, giving a combined effect of double centering on zero through the product $\hat{\Pi}_2\beta_{IV}$. In effect, shrinkage (achieved in this case by multiplication) helps to reduce variance, mirroring one of the broad ideas in Maasoumi (1978). Importantly, structural information is seen to be useful here where there is apparent overidentification even though the structural parameters themselves are unidentified and even though there are ostensibly more coefficients estimated in the pair $\left(\hat{\Pi}_2, \beta_{IV}\right)$ than there are in the simple unrestricted reduced form estimate $\hat{\pi}_1$.

The corresponding forecast mean square error (FMSE) values are

$$\mathbb{E}\left(\hat{y}_{1,n+1} - y_{1,n+1}\right)^2 = 1 + \frac{1}{n}, \text{ and } \mathbb{E}\left(\tilde{y}_{1,n+1} - y_{1,n+1}\right)^2 = 1 + \frac{m}{nK},$$

so that the restricted forecast also has a lower FMSE than the unrestricted forecast in this case.

The exact densities (25) and (26) do not depend on n. When we relax the assumption of Gaussianity in the error matrix V and use instead the CLT (17), we again obtain (25) and (26) but now as the limit distributions of $\hat{y}^P_{1,n+1}$ and $\tilde{y}^P_{1,n+1}$ as $n \to \infty$. Thus, all of the above stated properties of the two predictors hold asymptotically without Gaussianity.

A particularly interesting feature of the exact distribution (26) is that we may consider its limit as the number of instruments K (or degree of overidentification $K - m$) passes to infinity. This limit is most easily obtained by using the characteristic function of $\zeta = \tilde{y}^P_{1,n+1} = z'_{n+1}\left(\sqrt{n}\tilde{\pi}_2\right)$ that is given in the proof of the theorem in (37), viz.,

$$cf_\zeta\left(s\right) = e^{-\frac{s^2}{2}}\ {}_1F_1\left(\frac{K-m}{2}, \frac{K}{2}; \frac{s^2}{2}\right) = e^{-\frac{s^2}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j}{j!\left(\frac{K}{2}\right)_j}\left(\frac{s^2}{2}\right)^j. \tag{29}$$

This characteristic function is the exact finite sample and asymptotic (as $n \to \infty$) characteristic function of ζ. When $K \to \infty$ (under the assumption that $K < n \to \infty$) it is clear from (29)[6] that

$$\lim_{K\to\infty} cf_\zeta\left(s\right) = e^{-\frac{s^2}{2}} \sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{s^2}{2}\right)^j = 1.$$

It follows that the limit distribution of the restricted predictor $\tilde{y}^P_{1,n+1} = z'_{n+1}\left(\sqrt{n}\tilde{\pi}_2\right)$ has a point probability mass of unity at the origin. Hence, $\tilde{y}^P_{1,n+1} \to_p 0$ as $K \to \infty$. Correspondingly, the predictor variance $\mathbb{E}\left\{\left(\tilde{y}^P_{1,n+1}\right)^2\right\} = \frac{m}{K} \to 0$ as $K \to \infty$. Thus, increasing the number of (irrelevant) instruments has a dramatic shrinking effect on the restricted predictor, just as it does in fact for the distribution of β_{IV}.[7]

(b) The Weak IV Case

We next consider the weak IV case. We have the same predictors $\hat{y}_{1,n+1} = z'_{n+1}\hat{\pi}_1$ and $\tilde{y}_{1,n+1} = z'_{n+1}\tilde{\pi}_1 = z'_{n+1}\hat{\Pi}_2\beta_{IV}$, but now $[\pi_1, \Pi_2] = n^{-1/2}\left[\pi_1^*, \Pi_2^*\right]$ with $\pi_1^* = \Pi_2^*\beta$. The normalized unrestricted reduced form estimates satisfy $\sqrt{n}\left[\hat{\pi}_1, \hat{\Pi}_2\right] \sim_d N\left(\left[\pi_1^*, \Pi_2^*\right], I_{K(m+1)}\right) = N\left(\left[\Pi_2^*\beta, \Pi_2^*\right], I_{K(m+1)}\right)$ so that $\hat{y}^P_{1,n+1} =$

[6] The result follows by examining the expansion (29), letting $K \to \infty$, and noting that $\left(\frac{K-m}{2}\right)_j / \left(\frac{K}{2}\right)_j \to 1$ as $K \to \infty$ for all j. Alternatively, we may use the large-parameter asymptotic expansion ${}_1F_1\left(a, c; x\right) = e^x\left[1 + O\left(|c|^{-1}\right)\right]$ which holds when $c \to \infty$ and $c - a$ and x are bounded, as in the present case (see Erdélyi, 1953, p. 279).

[7] Recall from footnote 4 that the exact density (16) of β_{IV} in the irrelevant instrument case is the scaled multivariate t- distribution $t_q/q^{1/2}$, with $q = K - m + 1$ degrees of freedom, which collapses to the origin as $K \to \infty$.

$\sqrt{n}\hat{y}_{1,n+1} = \sqrt{n}z'_{n+1}\hat{\pi}_1 \sim N\left(z'\Pi_2^*\beta, 1\right)$. For the restricted reduced form estimator, we know that in this weak IV case with K fixed we have $\beta_{IV} \Rightarrow \zeta_\beta$ where ζ_β has the distribution given in (13) as $n \to \infty$ after making the replacements $\bar{\Pi}_{22} \mapsto n^{-1/2}\Pi_2^*$ and $m + L \mapsto K$ in the density. Then, as $n \to \infty$,

$$\begin{aligned} \tilde{y}_{1,n+1}^P &= \sqrt{n}\tilde{y}_{1,n+1} = \sqrt{n}z'_{n+1}\hat{\Pi}_2\beta_{IV} \\ &= z'\left(\Pi_2^* + C'V_2\right)\beta_{IV} \sim_a N\left(z'\Pi_2^*, I_m\right) \times \zeta_\beta, \end{aligned}$$

where β_{IV} can be written in component form as follows

$$\beta_{IV} = \left[\left(\Pi_2^{*\prime} + V_2'C\right)\left(\Pi_2^* + C'V_2\right)\right]^{-1}\left[\left(\Pi_2^{*\prime} + V_2'C\right)\left(\pi_1^* + C'v_1\right)\right].$$

Under either Gaussianity or asymptotically as $n \to \infty$, we have $C'V_2 \sim_d N\left(0, I_{Km}\right)$ and, as before, $C'V_2$ is independent of $C'v_1 \sim_d N\left(0, I_K\right)$.

Setting $W_\Pi = \left(\Pi_2^* + C'V_2\right) \sim_d N\left(\Pi_2^*, I_{Km}\right)$, we have the following representation, by virtue of initial conditioning on $C'V_2$,

$$\begin{aligned} \left(\Pi_2^* + C'V_2\right)\beta_{IV} &= W_\Pi\left(W_\Pi'W_\Pi\right)^{-1}\left[W_\Pi'\left(\Pi_2^*\beta + C'v_1\right)\right] \\ &= P_{W_\Pi}\Pi_2^*\beta + P_{W_\Pi}C'v_1 \sim_d \mathrm{MN}\left(P_{W_\Pi}\Pi_2^*\beta, P_{W_\Pi}\right), \end{aligned}$$

which involves a mean and covariance matrix normal mixture, for which we use the symbolism $\mathrm{MN}(\cdot,\cdot)$. Hence, the normalized predictor

$$\tilde{y}_{1,n+1}^P = \sqrt{n}z'_{n+1}\hat{\Pi}_2\beta_{IV} \sim_d \mathrm{MN}\left(z'P_{W_\Pi}\Pi_2^*\beta, z'P_{W_\Pi}z\right). \tag{30}$$

When $\Pi_2^* = 0$ this distribution reduces to $\mathrm{MN}\left(0, z'P_{C'V_2}z\right)$ as in the decomposition (27) used for the central case; and, when $\beta = 0$, we have $\tilde{y}_{1,n+1}^P \sim_d \mathrm{MN}\left(0, z'P_{W_\Pi}z\right)$. In both these cases we have a variance matrix normal mixture. Since P_{W_Π} is a projection matrix and W_Π has a full rank normal distribution which implies that $W_\Pi'z \neq 0$ with probability one, it follows that $z'P_{W_\Pi}z \in (0,1)$ *a.s.*. So, just as in the completely unidentified case, the variance of the normalized restricted predictor is smaller than the variance of the unrestricted predictor in the weak IV case when $\beta = 0$.

It is also possible to analyze the case of many weak instruments as $K \to \infty$. For example, suppose $z = i_K/\sqrt{K}$ where $i_K = (1, ..., 1)'$ is a K- vector with unity in each position, and assume that $K^{-1}i_K'\Pi_2^* \to_p 0$ and $K^{-1}\Pi_2^{*\prime}\Pi_2^* \to_p \Omega_\Pi > 0$ for some positive definite $m \times m$ matrix Ω_Π. These conditions are satisfied, for instance, when $\Pi_2^* = \left[\Pi_{21}^*, ..., \Pi_{2K}^*\right]'$ in which the component vectors $\Pi_{2k}^* \sim_{iid} \left(0, \Omega_\Pi\right)$ and are independent of V. Since $C'V_2 \sim_d N\left(0, I_{Km}\right)$ either exactly under Gaussianity or asymptotically using (17), we have $K^{-1}V_2'CC'V_2 \to_p I_m$, $K^{-1}V_2'C\Pi_2^* \to_p 0$, and $K^{-1}i_K'C'V_2 \to_p 0$. The quadratic form $z'P_{W_\Pi}z =$

$K^{-1}i_K' P_{W_\Pi} i_K$ then satisfies

$$
\begin{aligned}
z' P_{W_\Pi} z &= \frac{1}{K} i_K' \left(\Pi_2^* + C'V_2\right) \left[\left(\Pi_2^{*\prime} + V_2'C\right)\left(\Pi_2^* + C'V_2\right)\right]^{-1} \left(\Pi_2^{*\prime} + V_2'C\right) i_K \\
= \quad & \left(\frac{i_K'\Pi_2^*}{K} + \frac{i_K' C'V_2}{K}\right) \left[\left(\frac{\Pi_2^{*\prime}\Pi_2^*}{K} + \frac{\Pi_2^{*\prime}C'V_2}{K} + \frac{V_2'C\Pi_2^*}{K} + \frac{V_2'CC'V_2}{K}\right)\right]^{-1} \\
& \times \left(\frac{\Pi_2^{*\prime} i_K}{K} + \frac{V_2'C i_K}{K}\right) \\
\to_p \ & 0.
\end{aligned}
$$

Hence, we may expect the variance of the predictor $\tilde{y}_{1,n+1}^P$ to decrease as $K \to \infty$ in the many weak IV case, just as in the unidentified case.

The location variate in (30) is also of interest when $\beta \neq 0$. If we assume the conditions above that the weak reduced form parameters $\Pi_{2k}^* \sim_{iid} (0, \Omega_\Pi)$ and are independent of V, then $K^{-1/2} i_K' \Pi_2^* \Rightarrow \zeta_{\Pi_2} \sim_a N(0, \Omega_\Pi)$ and since $K^{-1/2} i_K' C'V_2 \Rightarrow \zeta_V \sim_d N(0, I_m)$, we find that

$$
\begin{aligned}
z' P_{W_\Pi} \Pi_2^* \beta &= \frac{1}{\sqrt{K}} i_K' \left(\Pi_2^* + C'V_2\right) \left[\left(\Pi_2^{*\prime} + V_2'C\right)\left(\Pi_2^* + C'V_2\right)\right]^{-1} \left(\Pi_2^{*\prime} + V_2'C\right) \Pi_2^* \beta \\
= \quad & \left(\frac{i_K'\Pi_2^*}{\sqrt{K}} + \frac{i_K' C'V_2}{\sqrt{K}}\right) \left[\left(\frac{\Pi_2^{*\prime}\Pi_2^*}{K} + \frac{\Pi_2^{*\prime}C'V_2}{K} + \frac{V_2'C\Pi_2^*}{K} + \frac{V_2'CC'V_2}{K}\right)\right]^{-1} \\
& \times \left(\frac{\Pi_2^{*\prime}\Pi_2^*}{K} + \frac{V_2'C\Pi_2^*}{K}\right) \beta \\
\Rightarrow \quad & \left(\zeta_{\Pi_2}' + \zeta_V'\right) \left(\Omega_\Pi + I_m\right)^{-1} \Omega_\Pi \beta,
\end{aligned}
$$

which shows that prediction bias can be expected in the many weak IV case when the structural coefficients $\beta \neq 0$. This bias is related to the fact that the IV estimate β_{IV} is inconsistent in this particular many weak IV case. Specifically, for given K the finite sample (or asymptotic distribution under (17) as $n \to \infty$) distribution is given by

$$
\begin{aligned}
\beta_{IV} &= \left[\left(\Pi_2^{*\prime} + V_2'C\right)\left(\Pi_2^* + C'V_2\right)\right]^{-1} \left[\left(\Pi_2^{*\prime} + V_2'C\right)\left(\pi_1^* + C'v_1\right)\right] \\
&= \left(W_\Pi' W_\Pi\right)^{-1} W_\Pi' \left(\pi_1^* + C'v_1\right) = \left(W_\Pi' W_\Pi\right)^{-1} W_\Pi' \left(\Pi_2^*\beta + C'v_1\right) \\
&\sim_d \mathrm{MN}\left(\left(W_\Pi' W_\Pi\right)^{-1} W_\Pi' \Pi_2^* \beta, \left(W_\Pi' W_\Pi\right)^{-1}\right).
\end{aligned}
$$

As above, we have

$$
K^{-1} W_\Pi' W_\Pi = \left(\frac{\Pi_2^{*\prime}\Pi_2^*}{K} + \frac{\Pi_2^{*\prime}C'V_2}{K} + \frac{V_2'C\Pi_2^*}{K} + \frac{V_2'CC'V_2}{K}\right) \to_p \Omega_\Pi + I_m,
$$

when $K \to \infty$, and so

$$
\begin{aligned}
\left(W_\Pi' W_\Pi\right)^{-1} W_\Pi' \Pi_2^* \beta &= \left(\frac{W_\Pi' W_\Pi}{K}\right)^{-1} \frac{W_\Pi' \Pi_2^* \beta}{K} \to_p \left(\Omega_\Pi + I_m\right)^{-1} \Omega_\Pi \beta \\
= \quad & \left[\Omega_\Pi^{-1} - \Omega_\Pi^{-1}\left(\Omega_\Pi^{-1} + I\right)^{-1} \Omega_\Pi^{-1}\right] \Omega_\Pi \beta = \beta - \left(\Omega_\Pi + I_m\right)^{-1} \beta.
\end{aligned}
$$

Hence, in this case,

$$\beta_{IV} \to_p \beta - (\Omega_\Pi + I_m)^{-1}\beta \tag{31}$$

and the IV estimator is inconsistent, a result that was obtained earlier in Chao and Swanson (2005, theorem 2.4(b)). While a limit distribution theory is possible, it is obviously of less interest due to the inconsistency. Observe that $\Pi_2^{*\prime}\Pi_2^* = O(K)$ here, which ensures that in Chao and Swanson's notation their $r_n = K$, and so $r_n/K \to 1$, thereby establishing the correspondence with their result. By contrast, the many weak IV case (iv) considered earlier has the concentration matrix expansion rate $\Pi_2^{*\prime}\Pi_2^* = O\left(n^{2\alpha}\right)$ with $K/n^{2\alpha} \to 0$, and β_{IV} is consistent at rate $O(n^\alpha)$ and asymptotically normal. Thus, consistency holds when the instrument set does not expand too fast in relation to information accumulated in the concentration matrix as $K \to \infty$. Again, a limit distribution theory is possible.

In the general noncentral weak IV case where $\Pi_2^* \neq 0$ and $\beta \neq 0$, the distribution of the predictor is the mean and variance matrix mixture of normals (30) in which the conditional mean vector $z' P_{W_\Pi} \Pi_2^* \beta$ is random. This substantially complicates the exact distribution theory and introduces a (random) conditional bias effect into the predictor that may not be eliminated asymptotically and may inflate variance. Although there may still be conditional variance reductions in this case because the conditional variance in (30) has the same value $z' P_{W_\Pi} z$, the overall effect of the random mean mixture on the properties of the restricted predictor $\tilde{y}_{1,n+1}^P$ is mixed. Some simulation evidence is presented next to provide further information on these properties.

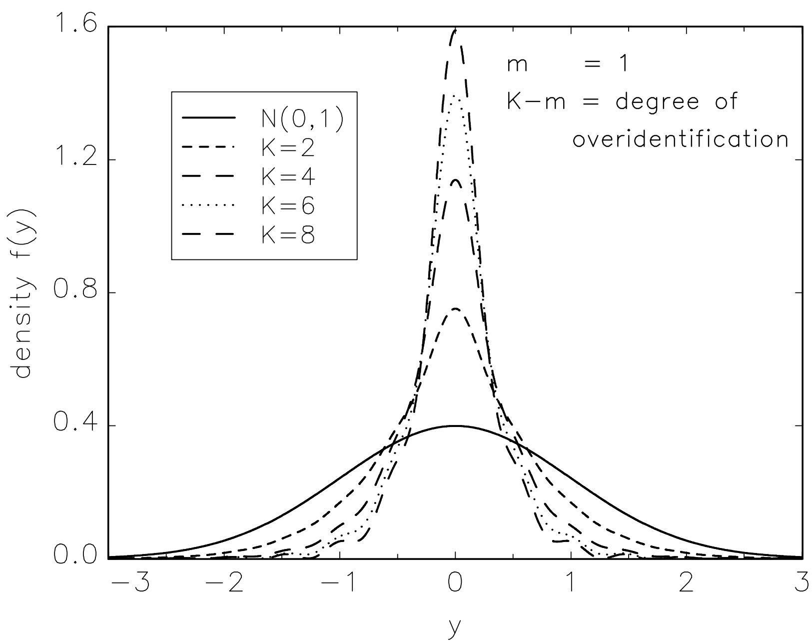

Fig. 1: Exact densities of reduced form and partially restricted reduced form

predictors in an unidentified structural model with $m+1=2$ endogenous variables and $K-m$ degrees of apparent overidentification in the equation.

3.2 Simulations

We first consider the unidentified case and produce numerical illustrations of the density results (25) and (26) given in the theorem. The IV (restricted) predictor density is readily computed directly from the analytic expression (26). The unrestricted predictor density is standard normal. Figures 1 and 2 show these densities for various values of the degree of overidentification $K-m$ and for $m=1$ and $m=3$. As is clear in both figures, the densities of the IV predictor are considerably more concentrated than that of the unrestricted predictor. Moreover, the concentration increases as $K-m$ increases. These numerical results confirm the analytic findings and show that the forecast variance reduction is particularly dramatic when $m=1$.

Figures 3 and 4 provide simulation-based kernel density estimates of the unrestricted and partially restricted reduced form (PRRF) predictors in the weak IV case. We report results for the two endogenous variable case (i.e., $m=1$ and there is a single right hand endogenous variable in the structural equation (22)), sample size $n=100$, and instrument numbers $K=3,11$ and 19. The K-vector Π_2^* has its elements drawn (once and for all) from the normal distribution $N\left(0, I_{Km}\right)$, $\Pi_2 = n^{-1/2}\Pi_2^*$, giving a conventional weak instrument matrix, and the forecast period vector $z_{n+1} = i_K/\sqrt{K}$. The structural coefficient has values $\beta=0$ in Figure 3, and $\beta=1$ in Figure 4. These differences in the values of the structural coefficient β turn out to be of substantial importance, as inspection of Figures 3-4 confirm.

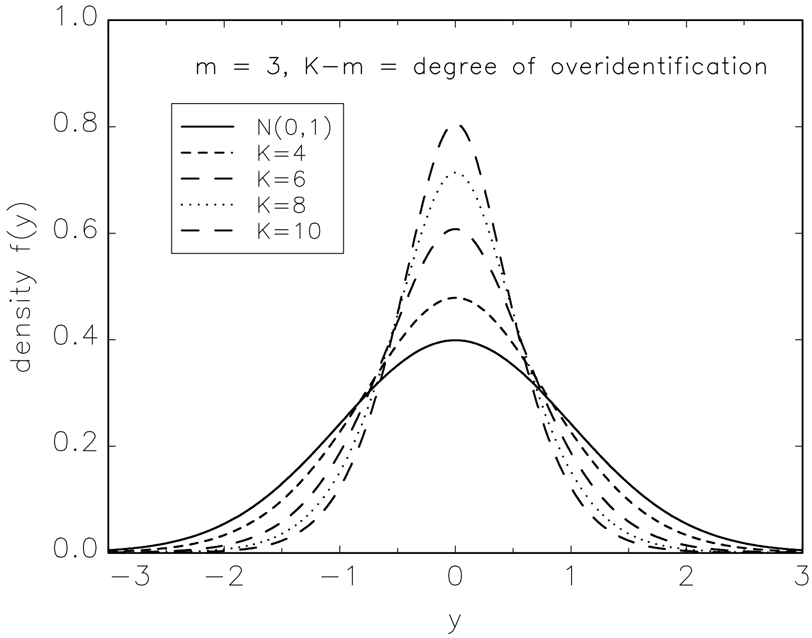

Fig. 2: Exact densities of reduced form and partially restricted reduced form predictors in an unidentified structural model with $m+1=4$ endogenous variables and $K-m$ degrees of apparent overidentification in the equation.

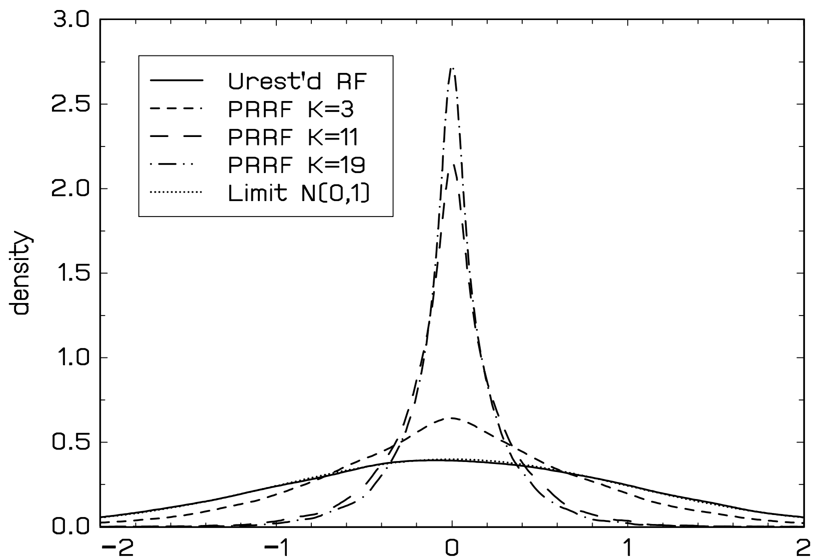

Fig. 3: Densities of the unrestricted and partially restricted reduced form predictors in the noncentral weak IV case when $\beta=0$. Simulations with $n=100$, $m=1$, $K=3,11,19$, $z_{n+1}=i_K/\sqrt{K}$,and $50,000$ replications.

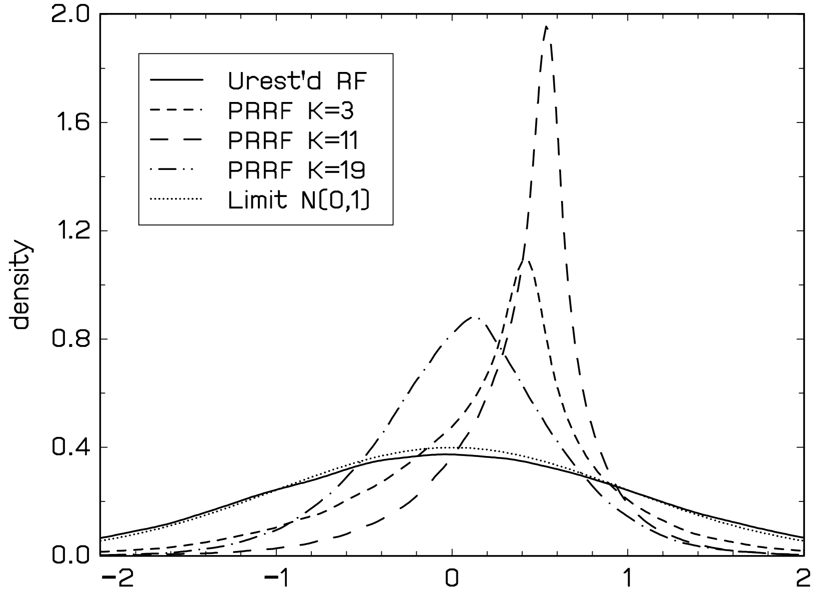

Fig. 4: Densities of the centred and scaled unrestricted reduced form predictor and the partially restricted reduced form predictor in the noncentral weak IV case when $\beta = 1$. Simulations are based on $n = 100$, $m = 1$, $K = 3, 11, 19$, $z_{n+1} = i_K/\sqrt{K}$,and $50,000$ replications.

First, it is clear from Figure 3 that when $\beta = 0$, the predictor distributions are all centred on the origin as, conditional on Π_2^*, the mean forecast period value is $\mathbb{E}\left(y_{1,n+1}\right) = z_{n+1}'\pi_1 = z_{n+1}'\Pi_2\beta = 0$. Second, the unrestricted predictor $\hat{y}_{1,n+1}^P = \sqrt{n}z_{n+1}'\hat{\pi}_1 \sim_d N\left(0,1\right)$ since $z_{n+1}'z_{n+1} = 1$, and it is apparent in the figure that the finite sample density matches the standard normal density. Third, the density of the PRRF predictor depends closely on the degree of overidentification $K-1$. Even for small degrees of overidentification like $K-1 = 2$, the PRRF predictor has a more concentrated density. As K increases, the density concentrates sharply, corroborating the analytic finding that the variance tends to zero as $K \to \infty$.

Figure 4 gives the densities for the centred and scaled versions of the predictors $\sqrt{n}\left(\hat{y}_{1,n+1} - y_{1,n+1}\right)$ and $\sqrt{n}\left(\tilde{y}_{1,n+1} - y_{1,n+1}\right)$ when $\beta = 1$ and the remaining parameters are as in Figure 3. Again, the distribution of the unrestricted predictor matches the asymptotic $N\left(0,1\right)$. The densities of the PRRF predictors show non-Gaussian features with peakedness, skewness, and location bias for all K, reflecting the mean and variance mixed normal form of the analytic distribution (30). In each case, the density is more concentrated than that of the unrestricted predictor, which corroborates the limit theory. The location bias is particularly noticeable for $K = 3$ and $K = 11$ and reflects the bias in the IV estimate (30) that is present for fixed K as $n \to \infty$. However, the density for $K = 19$ in Figure 4 shows that for larger K, the bias is reduced and the predictor distribution is better centered. This evidence matches the asymptotic theory that β_{IV} is consistent in the many weak instrument case, provided K does not increase too fast.[8]

4 Reflections on Exact Theory and Reduced Forms

At the time when Maasoumi's (1978) paper appeared, finite sample theory (including exact distribution theory) was the Rolls Royce of research areas in econometrics, a position it had occupied for nearly a decade. The avalanche of asymptotic theory that emerged in the 1980s appeared for a while to bury much of this literature. But the relentless search for generality in econometric methods never obscured the reality of the finite sample dependencies of econometric estimation and inferential methods that were one of the highlights of finite sample results. The recognition that asymptotics reproduce such dependencies in models that are only partly identified brought a revival of interest in exact distribution theory and analytic methods of approximation that capture the central features of those parameter dependencies.

This recognition, in its turn, put a premium on the search for inferential methods that assist in achieving some robustness to such dependencies. In structural equation models, the use of reduced form methods for inference on structural coefficients was noted by Chernozukov and Hansen (2008) who used the identifiability relation $\pi_1 = \Pi_2\beta$ to construct robust tests of hypotheses on the structural coefficients β using unrestricted reduced form estimates of π_1. Similar testing methods may be employed with partially restricted reduced form estimates and are currently under investigation.

The present paper uses this same correspondence but in a different manner and with a different focus to seek improvements in forecasting. The results reveal that such improvements are possible and can apply in the case where β is unidentified, thereby showing that information about the structural form may be useful even in the absence of effective instrumentation. These improvements are partly induced by shrinkage, just as those originally implemented by Maasoumi (1978) on system-wide reduced form estimates.

[8] For a detailed analysis of the bias properties of IV estimation, but not prediction, readers are referred to Chao and Swanson (2007), who consider various cases that allow for different expansion rates of K and n passing to infinity in IV estimation under weak instrumentation.

5 Appendix

The following lemma gives a matrix space integral that is useful in multivariate analysis. Part (a) is a minor extension of a standard result (Muirhead, 1982). We need this lemma and the characteristic function given in Part (b) in our main development.

Lemma A

(a) *If the $p \times p$ matrix X is positive semi-definite and $V_{q,p}$ is the Stiefel manifold of $q < p$ frames of $p \times 1$ orthonormal vectors*

$$\int_{V_{q,p}} \operatorname{etr}\left\{XH_1H_1'\right\}(dH_1) = {}_1F_1\left(\frac{q}{2};\frac{p}{2};X\right), \tag{32}$$

where the $p \times q$ matrix $H_1 \in V_{q,p}$ with $H_1'H_1 = I_p$, (dH_1) is the (normalized) Haar measure on $V_{q,p}$ so that $\int_{V_{q,p}}(dH_1) = 1$, and ${}_1F_1\left(\frac{q}{2};\frac{p}{2};X\right)$ is a matrix argument confluent hypergeometric function.

(b) *The characteristic function $cf_\zeta(s) = \mathbb{E}\left(e^{is\zeta}\right)$ of the random vector $\zeta \sim MN\left(0, z'\Upsilon\Upsilon'z\right)$, where $\Upsilon = \Xi\left(\Xi'\Xi\right)^{-1/2}$, Ξ is $K \times m$ with $K > m$ and distributed as matrix normal $\Xi \sim N\left(0, I_{mK}\right)$, and z is a fixed K- vector with $z'z = 1$, is given by*

$$cf_\zeta(s) = e^{-\frac{s^2}{2}}\,{}_1F_1\left(\frac{K-m}{2},\frac{K}{2};\frac{s^2}{2}\right) = e^{-\frac{s^2}{2}}\sum_{j=0}^{\infty}\frac{\left(\frac{K-m}{2}\right)_j}{j!\left(\frac{K_2}{2}\right)_j}\left(\frac{s^2}{2}\right)^j. \tag{33}$$

Proof of Lemma A

Part (a) Muirhead (1982, p.288 exercise 7.8) gives (32) for positive definite X. We show the result also holds when X is positive semi-definite. Set $H = [H_1, H_2] \in O(p)$, where H_2 is an orthogonal complement to H_1. Write $HH' = H_1H_1' + H_2H_2'$, so that $\operatorname{etr}\left\{XH_1H_1'\right\} = \operatorname{etr}\left\{XH_1H_1' + 0H_2H_2'\right\} = \operatorname{etr}\left\{XHYH'\right\}$ with

$$H_1H_1' = [H_1, H_2]\begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} H_1' \\ H_2' \end{bmatrix}, \text{ and } Y = \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}.$$

From Constantine (1963), James (1964) we have the zonal polynomial formula

$$C_\varphi(I_q) = c(\varphi)\,Z_\varphi(I_q)/1.3....(2q-1) = c(\varphi)\,2^{2j}\left(\frac{q}{2}\right)_\varphi\frac{j!}{(2j)!}, \tag{34}$$

where $c(\varphi)$ is a degree representation of the symmetric group (see formulae (23) - (25) of James (1961) and Muirhead (1982, p. 272)), $\varphi = (j_1, ..., j_q)$ is a partition of the integer j into not more than q parts, and

$(a)_\varphi = \prod_{i=1}^{q} \left(a - \frac{1}{2}(i-1)\right)_{j_i}$ is a generalized hypergeometric coefficient or forward factorial. Then, by standard multivariate methods (Muirhead, 1982, theorem 7.3.3. and lemma 7.5.7), utilizing (34), and noting that $C_\varphi(I_q) = 0$ if φ is a partition into more than q parts (Muirhead, 1982, p. 272), we obtain

$$\begin{aligned}
\int_{V_{q,p}} \operatorname{etr}\{XH_1H_1'\}(dH_1) &= \int_{O(p)} \operatorname{etr}\{XHYH'\}(dH) \\
&= {}_0F_0^{(p)}(X,Y) = \sum_{j=0}^{\infty}\sum_{\varphi} \frac{C_\varphi(X)\,C_\varphi(I_q)}{j!C_\varphi(I_p)} \\
&= \sum_{j=0}^{\infty}\sum_{\varphi} \frac{C_\varphi(X)\left[2^{2j}j!\left(\frac{q}{2}\right)_\varphi/(2j)!\right]}{j!\left[2^{2j}j!\left(\frac{p}{2}\right)_\varphi/(2j)!\right]} \\
&= \sum_{j=0}^{\infty}\sum_{\varphi} \frac{\left(\frac{q}{2}\right)_\varphi C_\varphi(X)}{j!\left(\frac{p}{2}\right)_\varphi} \\
&= {}_1F_1\left(\frac{q}{2};\frac{p}{2};X\right),
\end{aligned}$$

as required for (32) and this derivation holds for X positive semi-definite. The final line above uses the series representation of the matrix argument confluent hypergeometric function.

Part (b) If $\Xi \sim N(0, I_{Km})$ then the $K \times m$ matrix $\Upsilon = \Xi(\Xi'\Xi)^{-1/2} \in V_{m,K}$ $(K \geq m)$ is easily seen to be uniformly distributed on the Stiefel manifold $V_{m,K}$. Thus, we need to find the characteristic function of the random vector $\zeta \sim MN(0, z'\Upsilon\Upsilon'z)$ where Υ is uniform on $V_{m,K}$ and z is a fixed K- vector with $z'z = 1$. First observe that if $K = m$ then $\Upsilon = I_m$ and $\zeta \sim N(0,1)$ so that in this case $cf_\zeta(s) = \mathbb{E}\left(e^{is\zeta}\right) = e^{-\frac{s^2}{2}}$. In the general case with $K > m$ we need to resolve the following integral

$$cf_\zeta(s) = \mathbb{E}\left(\mathbb{E}\left(e^{is\zeta}|\Upsilon_m\right)\right) = \mathbb{E}\left(e^{-\frac{s^2}{2}z'\Upsilon\Upsilon'z}\right) = \int_{V_{K,m}} \operatorname{etr}\left\{-\frac{s^2}{2}\Upsilon\Upsilon'zz'\right\}(d\Upsilon),$$

where $(d\Upsilon)$ is the normalized invariant measure on the Stiefel manifold $V_{m,K}$. Construct the orthonormal matrix $H = [z, z_\perp] \in O(K)$, where $z_\perp$ is an orthogonal complement matrix for the vector z. Note that the measure $(d\Upsilon)$ is invariant under the transformation $\Upsilon \longmapsto H'\Upsilon = U \in V_{m,K}$. Then, defining the K- dimensional coordinate vector $e_1 = (1,0,...0)'$, using Part (a), and transforming the arguments of the ${}_1F_1$ function using the Kummer relation (Lebedev, 1972), we have

$$\begin{aligned}
cf_{\zeta}(s) &= \int_{V_{m,K}} \operatorname{etr}\left\{-\frac{s^2}{2}\Upsilon\Upsilon' zz'\right\}(d\Upsilon) = \int_{V_{m,K}} \operatorname{etr}\left\{-\frac{s^2}{2}U'H'zz'HU\right\}(dU) \\
&= \int_{V_{m,K}} \operatorname{etr}\left\{-\frac{s^2}{2}U'e_1e_1'U\right\}(dU) = {}_1F_1\left(\frac{m}{2},\frac{K}{2};-\frac{s^2}{2}e_1e_1'\right) \\
&= e^{-\frac{s^2}{2}}\ {}_1F_1\left(\frac{K-m}{2},\frac{K}{2};\frac{s^2}{2}e_1e_1'\right) = e^{-\frac{s^2}{2}}\ {}_1F_1\left(\frac{K-m}{2},\frac{K}{2};\frac{s^2}{2}\right) \\
&= e^{-\frac{s^2}{2}} \sum_{j=0}^{\infty}\frac{\left(\frac{K-m}{2}\right)_j}{j!\left(\frac{K_2}{2}\right)_j}\left(\frac{s^2}{2}\right)^j,
\end{aligned}$$

giving the result as stated. The penultimate line above follows because the matrix argument ${}_1F_1$ function reduces to the scalar argument ${}_1F_1$ function

$${}_1F_1\left(\frac{K-m}{2},\frac{K}{2};\frac{s^2}{2}e_1e_1'\right) ={}_1F_1\left(\frac{K-m}{2},\frac{K_2}{2};\frac{s^2}{2}\right),$$

since unity is the single non-zero eigenvalue of e_1e_1' and so zonal polynomials of $\frac{s^2}{2}e_1e_1'$ reduce to powers of $\frac{s^2}{2}$, giving $C_{\varphi}\left(\frac{s^2}{2}e_1e_1'\right) = \left(\frac{s^2}{2}\right)^j$ (c.f., Muirhead, 1982, corollary 7.2.4).

Proof of the Theorem

We seek to find the distribution of the predictor $\tilde{y}^P_{1,n+1} = z'_{n+1}(\sqrt{n}\tilde{\pi}_1) = z'_{n+1}\left(\sqrt{n}\hat{\Pi}_2\beta_{IV}\right)$. Suppose $K > m$. Under Gaussianity and after standardizing transformations, we have

$$\begin{aligned}
\sqrt{n}\tilde{\pi}_1 &= \sqrt{n}\left(Z'Z\right)^{-1}Z'Y_2 \times \left(V_2'CC'V_2\right)^{-1}V_2'CC'v_1 \\
&= C'V_2 \times \left(V_2'CC'V_2\right)^{-1}V_2'CC'v_1 \\
&= \left[\left(C'V_2\right)\left(V_2'CC'V_2\right)^{-1/2}\right]\left(V_2'CC'V_2\right)^{-1/2}\left(V_2'CC'v_1\right)
\end{aligned}$$

using $C' = \left(Z'Z\right)^{-1/2}Z'$ and $\sqrt{n}\left(Z'Z\right)^{-1/2} = I_K$. Define $\Xi = C'V_2 \sim_d N\left(0, I_{Km}\right)$, and it follows that $\Upsilon = \Xi\left(\Xi'\Xi\right)^{-1/2}$ is uniformly distributed on the Stiefel manifold $V_{m,K}$. The vector $\left(V_2'CC'V_2\right)^{-1/2}V_2'CC'v_1 \sim_d N\left(0, I_m\right)$ and is independent of Ξ and, hence, Υ. It follows that

$$\sqrt{n}\tilde{\pi}_1 \sim_d MN\left(0, C'V_2\left[V_2'CC'V_2\right]^{-1}V_2'C\right) = MN\left(0, \Upsilon\Upsilon'\right),$$

so that the exact distribution of $\sqrt{n}\tilde{\pi}_1$ is mixed normal with mixing variance matrix $\Upsilon\Upsilon' = \Xi\left(\Xi'\Xi\right)^{-1}\Xi'$ which projects onto $\mathcal{R}(\Xi)$. Then, setting $z_{n+1} = z$, we have $\tilde{y}^P_{1,n+1} = z'_{n+1}(\sqrt{n}\tilde{\pi}_1) \sim_d MN\left(0, z'\Upsilon\Upsilon'z\right) =: \zeta$. The required expression for the density is obtained by integrating out the mixed normal distribution with respect to Υ over the manifold $V_{m,K}$.

It is simplest to proceed by working with the characteristic function as follows. The forecast density of $\tilde{y}^P_{1,n+1}$ is, by inversion, $f_{\tilde{y}^P_{1,n+1}}(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isy} cf_{\zeta}(s)\, ds.$ Since $\zeta \sim_d MN(0, z'\Upsilon\Upsilon' z)$, its characteristic function is obtained directly from Lemma A as

$$cf_{\zeta}(s) = e^{-\frac{s^2}{2}} \, {}_1F_1\left(\frac{K-m}{2}, \frac{K}{2}; \frac{s^2}{2}\right) = e^{-\frac{s^2}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j}{j!\left(\frac{K}{2}\right)_j} \left(\frac{s^2}{2}\right)^j. \quad (35)$$

The required density follows by inversion. The series may be integrated term by term because the confluent hypergeometric series ${}_1F_1$ series is an entire function and uniformly convergent (e.g. Lebedev, 1972, p. 261). Proceeding, we obtain

$$\begin{aligned}
f_{\tilde{y}^P_{1,n+1}}(y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isy} e^{-\frac{s^2}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j}{j!\left(\frac{K}{2}\right)_j}\left(\frac{s^2}{2}\right)^j ds \\
&= \frac{1}{2\pi}\sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j \left(\frac{1}{2}\right)^j}{j!\left(\frac{K}{2}\right)_j} \int_{-\infty}^{\infty} e^{-isy} e^{-\frac{s^2}{2}} s^{2j} ds \\
&= \frac{1}{2\pi}\sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j \left(\frac{1}{2}\right)^j}{j!\left(\frac{K}{2}\right)_j} e^{-\frac{y^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(iy)^2}{2}} e^{-isy} e^{-\frac{s^2}{2}} s^{2j} ds \\
&= \frac{1}{\sqrt{2\pi}}\sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j \left(\frac{1}{2}\right)^j}{j!\left(\frac{K}{2}\right)_j} e^{-\frac{y^2}{2}} \int_{-\infty}^{\infty} \frac{e^{-\frac{(s+iy)^2}{2}} s^{2j}}{\sqrt{2\pi}} ds \\
&= \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j \left(\frac{1}{2}\right)^j}{j!\left(\frac{K}{2}\right)_j} \frac{2^j \Gamma\left(j+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \, {}_1F_1\left(-j; \frac{1}{2}; \frac{y^2}{2}\right) \\
&= \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\sum_{j=0}^{\infty} \frac{\left(\frac{K-m}{2}\right)_j \left(\frac{1}{2}\right)_j}{j!\left(\frac{K}{2}\right)_j} \, {}_1F_1\left(-j; \frac{1}{2}; \frac{y^2}{2}\right), \qquad (36)
\end{aligned}$$

where the fourth line is obtained by the recentering

$$\int_{-\infty}^{\infty} e^{-isy} e^{-\frac{s^2}{2}} s^{2j} ds = e^{-\frac{y^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(iy)^2}{2}} e^{-isy} e^{-\frac{s^2}{2}} s^{2j} ds = e^{-\frac{y^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(s+iy)^2}{2}} s^{2j} ds,$$

and the penultimate line by the even moment formula for the normal distribution. In particular, if $X \sim_d N(\mu, \sigma^2)$ we have

$$\mathbb{E}\left(X^{2j}\right) = \sigma^{2j} 2^j \frac{\Gamma\left(j+\frac{1}{2}\right)}{\sqrt{\pi}} \, {}_1F_1\left(-j, \frac{1}{2}, -\frac{\mu^2}{2\sigma^2}\right),$$

and, setting $\mu = iy$ and $\sigma = 1$, we have

$$\int_{-\infty}^{\infty} \frac{e^{-\frac{(s+iy)^2}{2}} s^{2j}}{\sqrt{2\pi}} ds = \frac{2^j \Gamma\left(j+\frac{1}{2}\right)}{\sqrt{\pi}} \, {}_1F_1\left(-j, \frac{1}{2}, \frac{y^2}{2}\right) = 2^j \left(\frac{1}{2}\right)_j \, {}_1F_1\left(-j, \frac{1}{2}, \frac{y^2}{2}\right).$$

The $K=m$ case is obtained from (36) by noting that the series truncates at the first term giving $f_{\tilde{y}^P_{1,n+1}}(y)=e^{-\frac{y^2}{2}}/\sqrt{2\pi}$. In fact, when $K=m$, the exact densities of the two predictors are the same and so $\tilde{y}^P_{1,n+1}=\hat{y}^P_{1,n+1}$, thereby giving (25) as a special case of (26).

Proof of the Corollary

To calculate moments, we first employ the characteristic function of the normalized predictor $\zeta=\tilde{y}^P_{1,n+1}=z'_{n+1}(\sqrt{n}\tilde{\pi}_2)\sim_d MN(0,z'\Upsilon\Upsilon'z)$, given earlier in (35), viz.,

$$cf_\zeta(s)=e^{-\frac{s^2}{2}}\,{}_1F_1\left(\frac{K-m}{2},\frac{K}{2};\frac{s^2}{2}\right)=e^{-\frac{s^2}{2}}\sum_{j=0}^{\infty}\frac{\left(\frac{K-m}{2}\right)_j}{j!\left(\frac{K}{2}\right)_j}\left(\frac{s^2}{2}\right)^j. \tag{37}$$

The function $cf_\zeta(s)$ is analytic because both $e^{-\frac{s^2}{2}}$ and ${}_1F_1\left(\frac{K-m}{2},\frac{K}{2};\frac{s^2}{2}\right)$ are entire functions. So, all moments of ζ exist and are finite. The series representation of (35) produces the Taylor expansion of $cf_\zeta(s)$, viz.,

$$\begin{aligned}cf_\zeta(s) &= e^{-\frac{s^2}{2}}\left(1+\frac{\frac{K-m}{2}}{\frac{K}{2}}\frac{s^2}{2}+\frac{\left(\frac{K-m}{2}\right)\left(\frac{K-m+2}{2}\right)}{2\left(\frac{K}{2}\right)\left(\frac{K}{2}\right)}\left(\frac{s^2}{2}\right)^2+O\left(s^6\right)\right)\\ &= 1+\frac{s^2}{2}\left(\frac{K-m}{K}-1\right)+O\left(s^4\right),\end{aligned}$$

so that $cf''_\zeta(0)=-\frac{m}{K}$. Hence $\mathbb{E}\left(\tilde{y}^P_{n+1}\right)^2=\frac{m}{K}$. Since $\tilde{y}^P_{1,n+1}=z'_{n+1}(\sqrt{n}\tilde{\pi}_1)=\sqrt{n}\tilde{y}_{1,n+1}$, it follows that $\mathbb{E}\left(\tilde{y}_{1,n+1}-y_{1,n+1}\right)^2=1+\frac{m}{nK}$, giving the forecast mean square error in the general case.

6 References

Breiman, L. (1996), "Bagging predictors," *Machine Learning*, 36, 105-139.

Chao, J. C. and N. R. Swanson (2005). "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673-1692.

Chao, J. and N. R. Swanson (2007). "Alternative Approximations of the Bias and MSE of the IV Estimator Under Weak Identification with Application to Bias Correction", *Journal of Econometrics,* 137, 515-555

Chernozukov V. and C. Hansen (2008). "The reduced form: a simple approach to inference with weak instruments", *Economics Letters*, 100, 68-71.

Constantine, A. G. (1963). "Some noncentral distribution problems in multivariate analysis," *Annals of Mathematical Statistics* 34, 1270–1285.

Doan, T., R.B. Litterman and C. Sims (1984). "Forecasting and conditional projections using realistic prior distributions," *Econometrics Reviews* 3, 1–100.

Erdélyi, A. (1953). *Higher Transcendental Functions*, Volume 1. Krieger: Malabar.

Fan, Y. and A. Ullah (1999). "Asymptotic Normality of a Combined Regression Estimator," *Journal of Multivariate Analysis,* 71, 191-240.

Fiorio, C. V., V. Hajivassiliou and P. C. B. Phillips, (2010). "Bimodal t-ratios: The Impact of Thick Tails on Inference", *Econometrics Journal*, 13, 271-289.

Hansen, B.E. (2007) "Least squares model averaging," *Econometrica*, 75, 1175–1189.

Hansen, B. E. (2009). "Averaging estimators for regressions with a possible structural break," *Econometric Theory*, 25, 1498-1514.

Hansen, B. E. (2014). "Model Averaging, Asymptotic Risk, and Regressor Groups," *Quantitative Economics*, (2014) 5, 495-530.

Iglesias, E. M. and G. D. A. Phillips (2012). "Almost Unbiased Estimation in Simultaneous Equations Models with Strong and/or Weak Instruments," *Journal of Business and Economic Statistics*, 30, 505-520

James, A. T. (1961). "Zonal polynomials of the real positive definite symmetric matrices", *Annals of Mathematics*, 74, 456-469.

James, A. T. (1964). "Distribution of matrix variates and latent roots derived from normal samples," *Annals of Mathematical Statistics* 35, 475–501.

Kakwani, N. C. and R. H. Court (1972). "Reduced Form Coefficient Estimation and Forecasting from a Simultaneous Equation Model", *Australian Journal of Statistics,* 14, 143-160.

Knight, J. L. (1977). "On the Existence of Moments of the Partially Restricted Reduced Form Estimators from a Simultaneous Equation Model", *Journal of Econometrics,* 5, 315-321.

Kotlyarova, Y. and V. Zinde-Walsh (2006). "Non- and semi-parametric estimation in models with unknown smoothness, *Economics Letters,* 93, 379-386.

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

Liao, Z. and P. C. B. Phillips (2015). "Automated Estimation of Vector Error Correction Models", *Econometric Theory*, 31, 584-646.

Maasoumi, E. (1978). "A Modified Stein-like Estimator for the Reduced Form Coefficients of Simultaneous Equations," *Econometrica,* 46, 695-703.

Malinvaud, E. (1966). *The Statistical Methods of Econometrics.* Amsterdam: North Holland.

McCarthy, M. D. (1972). "A Note on the Forecasting Properties of 2SLS Restricted Reduced Forms," *International Economic Review*, 13, 757-761.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Phillips, P. C. B. (1980). "The exact finite sample density of instrumental variable estimators in an equation with $n+1$ endogenous variables," *Econometrica* 48:4, 861–878.

Phillips, P. C. B. (1983). "Exact small sample theory in the simultaneous equations model," Chapter 8 and pp. 449–516 in M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics.* Amsterdam: North–Holland.

Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory* 5, 181–240.

Phillips, P. C. B. (1996). "Econometric model determination", *Econometrica*, 64, 763-812.

Phillips, P. C. B. (2006). "A remark on bimodality and weak instrumentation in structural equation estimation", *Econometric Theory*, 22, pp. 958 – 984.

Phillips, P. C. B. (2009). "Exact distribution theory in structural estimation with an Identity," *Econometric Theory*, 2009, pp. 958 – 984.

Phillips, P. C. B. & V. A. Hajivassiliou (1987). "Bimodal t Ratios," Cowles Foundation Discussion paper #842.

Sargan, J. D. (1958). "The estimation of economic relationships using instrumental variables," *Econometrica* 26, 393–415.

Sargan, J. D. (1959). "The estimation of relationships with autocorrelated residuals by the use of instrumental variables," Journal of the Royal Statistical Society, Series B, 21, 91-105.

Sargan, J.D. (1973). "The Tails of the FIML Estimates of the Reduced Form Coefficients," Mimeographed, London School of Economics, 1973.

Sargan, J.D. (1976/1988)," The existence of the moments of estimated reduced form coefficients", mimeo LSE, later published in Sargan, J. D. *Contributions to Econometrics, Vol. 2.* (1988) (Ed. E. Maasoumi), Cambridge: Cambridge University Press.

Sargan, J. D. (1988a). *Lectures on Advanced Econometric Theory*, (Ed. M. Desai), Oxford: Blackwell.

Sargan, J. D. (1988b). *Contributions to Econometrics, Vols. 1 and 2*. (Ed. E. Maasoumi), Cambridge: Cambridge University Press

Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65, 557-586.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso" *Journal of the Royal Statistical Society B*, 58, 267-288.

Winkelbauer, A. (2014). "Moments and absolute moments of the normal distribution", arXiv:1209.4340v2 [math.ST].

Zou, H., (2006). "The adaptive lasso and its oracle properties", *Journal of the American Statistical Association*, 101, 1418-1429.