Federal Reserve Bank of New York
Staff Reports

Generalized Canonical Regression

Arturo Estrella

Staff Report no. 288
June 2007

Generalized Canonical Regression
Arturo Estrella
Federal Reserve Bank of New York Staff Reports, no. 288
June 2007
JEL classification: C51, C22

Abstract

This paper introduces a generalized approach to canonical regression, in which a set of jointly dependent variables enters the left-hand side of the equation as a linear combination, formally like the linear combination of regressors in the right-hand side of the equation. Natural applications occur when the dependent variable is the sum of components that may optimally receive unequal weights or in time series models in which the appropriate timing of the dependent variable is not known a priori. The paper derives a quasi-maximum likelihood estimator as well as its asymptotic distribution and provides illustrative applications.

Key words: linear regression, time series, canonical correlations

---

Estrella: Federal Reserve Bank of New York (e-mail: arturo.estrella@ny.frb.org). The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

1. Introduction

1.1 Motivation

Linear regression assumes that there is a single dependent variable, which takes on a unique role within the equation. Frequently, however, the dependent variable is a sum of components that implicitly receive equal weights, or weights that are specified a priori in arbitrary fashion. For example, the dependent variable may consist of average GDP growth over four quarters, with equal weight assigned to each quarter. Alternatively, the dependent variable may be an accounting aggregate, such as total bank loans or total bank capital, which is the sum of several well-defined components.

However, some of those components may be more important than others in the equation, and some may not play a systematic role at all, contributing only noise. If some individual components are of special interest, they may be used as single dependent variables in separate regressions. However, in that case there is no unified basis for comparing the individual results.

In canonical regression, a set of jointly dependent variables enters the left hand side of the equation as a linear combination, formally like the linear combination of regressors in the right hand side of the equation. Canonical regression then determines simultaneously the best predictive linear combination on the right hand side and the most predictable linear combination on the left hand side of the equation.

In the traditional approach, canonical regression is an outgrowth of canonical correlation and is applied to zero-mean or demeaned variables. The analysis is generally based on least squares methods and does not consider explicitly either the stochastic properties of the variables or the stochastic distribution of coefficient estimates. Examples of empirical applications in the literature are relatively few.

This paper takes a generalized approach to canonical regression, allowing for variables with nonzero means and general distributions, including possibly a constant term. More importantly, the analysis focuses on the asymptotic distribution of coefficient estimates and on statistical inference with regard to the coefficients. Generalized canonical regression is interpreted as the conditional expectation of a linear combination of jointly dependent variables, and a quasi-maximum likelihood estimator is constructed using relatively weak assumptions about the underlying data-generating process.

Using the asymptotic distribution, all coefficients in a generalized canonical regression may be subjected to the standard variety of statistical tests, including individual significance tests and tests of the equal-weight restrictions implicit in the use of a sum or average as a single dependent variable. The technique may be applied to structural or non-structural models.

1.2 Related earlier literature

Canonical regression was developed by Bartlett (1938) as an extension of the canonical correlation analysis of Hotelling (1935, 1936). Whereas canonical correlation analysis focuses on correlations between linear combinations of two sets of variables, canonical regression deals with the estimation of a regression equation that corresponds to the largest, or "first," canonical correlation.

Although the term "canonical regression" does not appear explicitly in this early literature, it is used later by Tintner (1950) and Bartlett (1951) in reference to Bartlett (1938). Waugh (1942) used the term "regression" in the context of canonical correlation analysis, but did not present a regression equation as such. Rather, he considered the coefficients of the pair of canonical variates, or linear combinations, corresponding to the first canonical correlation. These coefficients were scaled separately and did not conform to a single regression equation. A similar approach is taken in Tintner (1946) and more recently in Margulis (1998).

Vinod (1969), Gyimah-Brempong and Gyapong (1991), and Ruggiero (1998) return to the canonical regression form of Bartlett (1938). In each paper, the relationship between the first pair of canonical variates is given explicitly in a single regression equation, with conformably scaled point estimates of the coefficients.

The glaring omission in traditional canonical regression analysis is some notion of the stochastic distribution of coefficient estimates, and therefore the ability to perform inference with regard to these estimates. Although the computation of the coefficients is straightforward, computation of their covariance matrix has been seen as more of a challenge. Trippi (1977) is exceptional in giving standard errors for the estimated coefficients of a canonical regression, but does not indicate the method used to compute them.

Recently, Anderson (1999, 2002) provided a methodology for calculating asymptotic variances and covariances for all coefficient estimates in standard canonical correlation analysis, under the assumption that the data are generated by a structural linear model with normally-distributed variables and iid disturbances. If these conditions hold, the coefficients of the first canonical pair correspond to those in a canonical regression, up to scaling, and the asymptotic distribution may be used for statistical inference about the coefficients. Anderson (1999, p. 11) gives explicit expressions in the case when variables are demeaned and $K = J$. The asymptotic covariances of the first set of canonical coefficients are functions of all the estimated canonical coefficients and correlations, not just the first.

The generalized canonical regression approach of the present paper differs from standard canonical regression in two principal respects. First, the regression equation is expressed directly and conformably in terms of linear functions of jointly dependent variables and regressors, with fairly non-restrictive distributional assumptions. As noted earlier, the variables may or may not be zero-mean and the equation may or may not contain a constant term.

The second and more important difference is that the stochastic properties of the parameter estimates are explicitly considered, and their asymptotic distribution is derived under fairly general assumptions. If the variables are normally distributed and the errors are independent, parameters are computed by maximum likelihood estimation. If not, parameters may be estimated consistently by a quasi-maximum likelihood estimator, whose asymptotic distribution is calculated. Inference with regard to parameters is then straightforward. Moreover, generalized canonical regression allows for the computation of the asymptotic distribution based only on the first canonical relationship.

Generalized canonical regression has a superficial similarity to a single equation in a simultaneous equations model, in that several jointly dependent variables are linearly related to a set of regressors.[1] An important difference between the two formulations, however, is that the coefficients in a simultaneous equation model are assumed to be structural and identifiable in general only with reference to the whole system, whereas generalized canonical regression is a conditional expectation in which the coefficients are not necessarily structural and do not require information outside the equation for identification and estimation.

The method of alternating conditional expectations (ACE) of Breiman and Friedman (1985) is similar in that it derives both an optimal predictor and an optimal predicted variable. Like generalized canonical regression, the resulting relationship maximizes the R^2 of the equation among all admissible equations. However, the dependent variable in ACE is a nonlinear function of a single dependent variable, rather than a linear combination of several jointly dependent variables, as in canonical regression.

[1] Regressors correspond heuristically to exogenous variables in simultaneous equations. See, e.g., Fisher (1966). Note that Hooper (1959) proposes canonical correlation (not regression) analysis as a means of assessing the overall fit of simultaneous equation models.

It should be noted that generalized canonical regression is not at all like models with lagged dependent variables, including ARMA models as in Box and Jenkins (1970) and dynamic regression models estimated by maximum likelihood as in Engle (1980). Lagged dependent variables are subject to orthogonality conditions that do not apply to the jointly dependent variables in generalized canonical regression. More broadly, treating any of the jointly dependent variables as regressors tends to produce very different results.

2. Model formulation

2.1 Definition

Generalized canonical regression takes the form

$$y_t{}'a = x_t{}'b + u_t \tag{1}$$

where for observations $t = 1,\ldots,n$, y_t is a vector of *J* jointly dependent variables, x_t is vector of *K* explanatory variables or regressors, u_t is a scalar disturbance, and *a* and *b* are vectors of *J* and *K* coefficients, respectively.[2] In standard linear regression, $J = a = 1$ and the single variable y_t has a unique role in the equation. In generalized canonical regression, all the *J* dependent variables in the vector y_t are treated analogously.

As in standard regression, $x_t{}'b$ represents the best predictor of the dependent variable. In this case, however, the dependent variable $y_t{}'a$ is the most predictable linear combination of the jointly dependent variables y_t. "Most predictable" may be defined in a least squares sense as minimizing the sum of square residuals in proportion to the variance of the linear combination,

[2] The terms "jointly dependent variables" and "explanatory variables" are used in the sense of White (1996, Definition 4.3).

$\sum_t u_t^2 / \sum_t \left(a' (y_t - \bar{y}) \right)^2$, or in the sense of maximizing the likelihood of the model among all linear combinations of y_t with a given variance.[3] The regression may be viewed as a conditional expectation $E\left({y_t}' a \mid x_t \right) = {x_t}' b$ or, alternatively, $E(u_t \mid x_t) = 0$.

Some geometric intuition may be obtained by considering equation (1) as defining a hyperplane $P_c = \left\{ y_t \in \mathbb{R}^J \mid {y_t}' a = c \right\}$ in the space of jointly dependent variables, where $c = {x_t}' b + u_t$. For the purposes of the model, only movement in the direction orthogonal to the plane is important, that is, movement along the gradient $\nabla({y_t}' a) = a$. Movement within P_c has absolutely no effect in the model.

More precisely, any given change in the vector y_t may be decomposed as

$$\Delta y_t = \frac{(\Delta y_t)' a}{a' a} a + r \,, \tag{2}$$

where the first term represents movement along a and the remainder r is orthogonal to a ($r' a = 0$). A change $\Delta x_{\cdot k}$ in one of the regressors, holding the others fixed, produces a translation of the plane by a magnitude $\Delta({y_t}' a) / |a| = (\Delta y_t)' a / |a| = b_k \Delta x_{\cdot k} / |a|$ in the direction $a / |a|$. Any further movement in y_t is orthogonal to a and does not figure in the model. The coordinates of a are thus indicative of the relative importance of the respective jointly dependent variables in the equation.

[3] The "most predictable" criterion may be traced back to Hotelling (1935).

2.2 Maximum likelihood estimator (MLE)

The generalized canonical regression model (1) may be expressed in matrix form by stacking observations in the standard way

$$Ya = Xb + u\,, \tag{3}$$

where we assume that *X* and *Y* are of full rank. If $u \sim N(0,\sigma^2 I)$, the log likelihood function of the model is

$$\log L = \sum_{t=1}^{n} \log \phi(u_t) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2} u'u\,, \tag{4}$$

where $\phi(\cdot)$ is the standard normal density function and σ^2 is the variance of u_t. To estimate the model, we maximize this expression, given the variance of $y_t{}'a$, to obtain a maximum likelihood estimate of the parameter vector $\theta = \left(a', b', \sigma^2\right)'$.

For simplicity and without loss of generality, we assume that $y_t{}'a$ has unit variance. This condition is necessary to avoid the trivial solution $\theta = 0$ and to identify a unique nonzero solution. It is clear from (1) and (4) that if $\theta = \left(a', b', \sigma^2\right)'$ is a solution, so is $\theta_\lambda = \left(\lambda a', \lambda b', \lambda^2 \sigma^2\right)'$ for any real $\lambda \neq 0$. The unit variance condition fixes the absolute scale of θ.[4] The sign of (a,b) is arbitrary but may be set, for instance, with regard to the sign of a single coefficient or a linear function of coefficients.

[4] In fact, if θ is the solution obtained under the unit variance condition, we may rescale the results by using θ_λ for any $\lambda \neq 0$. This rescaled solution is equivalent to a constraint that the variance of $y_t{}'a$ is λ^2. Inference about parameter significance or relative magnitudes is not affected.

2.3 Quasi-maximum likelihood estimator (QMLE)

Suppose now that the exact distribution of (x_t, y_t) is unknown, but that we are most interested in the conditional expectation

$$E\left(y_t{}'a \mid x_t\right) = x_t{}'b\,. \tag{5}$$

In that case, we can use the log likelihood function (4) to construct a quasi maximum likelihood estimate of the parameter vector θ, subject to the unit variance constraint on $y_t{}'a$, which under certain conditions will produce consistent estimates of the parameters *a* and *b*. The following set of assumptions may be used to prove consistency of the QMLE.

Assumption Set 1:

(i) The observed data are a realization of the stochastic process $\left\{(X_t', Y_t')' : \Omega \to \mathbb{R}^{K+J}, K, J \in \mathbb{N}, t = 1,2,\ldots\right\}$ on a complete probability space $(\Omega, \mathcal{F}, P_0)$, where $\Omega = \mathbb{R}^{(K+J)\infty}$ and $\mathcal{F} = \mathcal{B}\left(\mathbb{R}^{(K+J)\infty}\right)$.

(ii) For any nonzero vector $a \in \mathbb{R}^J$, the conditional mean of $a'Y_t$ given X_t is a collection of sequences $\mathcal{M} = \{\mu(\theta) : \theta \in \Theta\}, \quad \mu(\theta) = \left\{\mu_t(\cdot, \theta) : \mathbb{R}^{K+J} \to \mathbb{R}\right\}$, where $\theta \in \mathbb{R}^{K+J}$ includes the elements of *a*, Θ is a compact subset of $\mathbb{R}^{K+J}$, $\mu_t(\cdot, \theta)$ is measurable–$\mathcal{B}\left(\mathbb{R}^{K+J}\right)$, and $\mu_t(X_t, \cdot)$ is continuous on Θ, a. s.–P_0, $t = 1,2,\ldots$ for a given choice of regressors $\{X_t\}$.

(iii) $E|Y_t| < \infty$ and there is a $\theta_0 \in \Theta$ such that $\mu_t(X_t, \theta_0) = E(a'Y_t \mid X_t)$ a. s.–P_0, $t = 1,2,\ldots$

(iv) For each $\theta \in \Theta$, $E \log f(X_t, Y_t, \theta)$ exists and is finite, $t = 1,2,\ldots$

(v) $E \log f(X_t, Y_t, \cdot)$ is continuous on Θ, $t = 1,2,\ldots$

(vi) $\{\log f(X_t, Y_t, \theta)\}$ obeys the strong (weak) uniform law of large numbers.

(vii) The sequence $\left\{\frac{1}{n}\sum_{t=1}^{n} \log f(X_t, Y_t, \theta)\right\}$ has identifiably unique maximizers $\left\{\theta^*_{(n)}\right\}$ subject to the condition that $a'Y_t$ has unit variance.

Using the foregoing assumptions and definitions, we can state the following result, which is analogous to results in White (1996, Section 5).

Proposition 1. If Assumption Set 1 holds, the constrained QMLE derived from $\sum_t \log f(X_t, Y_t, \theta) = \log L$ as in (4) is a consistent estimator of the parameters θ, even if *u* is not normally distributed and iid.[5]

Assume for the moment that the variables (x_t, y_t) have zero means and nonsingular variance covariance matrix. These assumptions are relaxed in Section 2.3.4 below. To compute the QMLE (or MLE) of the generalized canonical regression model, we maximize the likelihood function (4) subject to the unit variance constraint, which may be expressed as

$$a'Y'Ya = n. \tag{6}$$

To do so, form the Lagrangian

$$\mathcal{L} = \log L - \mu\left(a'Y'Ya - n\right). \tag{7}$$

and calculate the QMLE $(\tilde{\theta}, \tilde{\mu})$ as

[5] Proofs of propositions are sketched out in the appendix.

$$(\tilde{\theta}, \tilde{\mu}) = \arg\min \mathcal{L}(\theta, \mu). \tag{8}$$

2.3.1 First order conditions

Since the particular likelihood function and constraint used here are continuous and twice differentiable, the constrained optimization problem may be solved by solving first order conditions and verifying that second order conditions are met.

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial \theta} = \left[-\sigma^{-2} Y'u - 2\mu Y'Ya \quad \sigma^{-2} X'u \quad -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} u'u \right]' = 0 \tag{9}$$

as well as the constraint in (6). The second and third blocks of conditions in (9) are fairly standard, mutatis mutandis. The third condition implies that

$$\tilde{\sigma}^2 = \tilde{u}'\tilde{u} / n, \tag{10}$$

where $\tilde{u} = Y\tilde{a} - X\tilde{b}$, which is a standard maximum likelihood result. The second block may be expressed as

$$\tilde{b} = \left(X'X \right)^{-1} X'Y\tilde{a}, \tag{11}$$

which parallels the standard linear regression result for the vector of regressor coefficients, with the linear combination $Y\tilde{a}$ taking the place of the usual single dependent variable. Clearly, the generalized canonical framework has no significant consequences for the estimation of the right hand side of the equation.

In contrast, the first condition represents a clear departure from standard regression:

$$\left(1 + 2\tilde{\mu}\tilde{\sigma}^2 \right) Y'Y\tilde{a} = Y'X\tilde{b} = Y'X \left(X'X \right)^{-1} X'Y\tilde{a}, \tag{12}$$

where the second equality makes use of (11). Rewriting as

$$\left[\left(1 + 2\tilde{\mu}\tilde{\sigma}^2 \right) I - \left(Y'Y \right)^{-1} Y'X \left(X'X \right)^{-1} X'Y \right] \tilde{a} = 0, \tag{13}$$

we see that $1+2\tilde{\mu}\tilde{\sigma}^2$ is an eigenvalue of $\left(Y'Y\right)^{-1}Y'X\left(X'X\right)^{-1}X'Y$. The solution is obtained by setting $1+2\tilde{\mu}\tilde{\sigma}^2 = R^2 = r_1$, where r_1 is the largest eigenvalue of $\left(Y'Y\right)^{-1}Y'X\left(X'X\right)^{-1}X'Y$. The estimate $\tilde{a}$ is the corresponding eigenvector, scaled to satisfy constraint (6).

As mentioned earlier, the quadratic variance constraint does not determine the sign of $(\tilde{a},\tilde{b})$. This indeterminacy is not a problem in general, since it does not affect the signs of ratios of the form $\tilde{a}_j/\tilde{b}_k$. However, if one regressor, say $x_{\cdot k}$, is of particular interest, it may be convenient to choose the sign of $\tilde{b}_k$ and adjust the signs of $\tilde{a}$ and $\tilde{b}$ accordingly. To summarize, we have the following.

Proposition 2. The QMLE of *a* is the eigenvector corresponding to the largest eigenvalue of $\left(Y'Y\right)^{-1}Y'X\left(X'X\right)^{-1}X'Y$, scaled to satisfy (6) and a sign constraint. The QMLEs of *b* and σ^2 are obtained more conventionally from expressions (11) and (10). In addition, $\tilde{\sigma}^2 = 1-R^2$ and $\tilde{\mu} = -1/2$.

This proposition may be used to clarify the relationship between generalized canonical regression and canonical correlations. If, as assumed thus far, the variables x_t and y_t have zero means, r_1 is the first squared canonical correlation between x_t and y_t, and the QMLE $\tilde{a}$ is proportional to the vector of coefficients of the first canonical component of y_t with respect to x_t.[6] These equivalences may not hold if the variables have non-zero means, as seen in Section 2.3.4. Moreover, the scaling conventions in canonical correlation analysis

[6] See, for instance, Theil (1971, Section 7.4).

are typically different; the coefficients of y_t and x_t are scaled independently and are not in general conformable in a regression setting.

A final note with regard to the first order conditions highlights an important feature of generalized canonical regression. In contrast to standard regression, in which all variables but one are orthogonal to the residual, the first block of (9) shows that none of the jointly dependent variables are orthogonal to the residual. Equation (6) and $\tilde{a}'Y'X\tilde{b} = nR^2$ (see proof of Proposition 2) imply that $\tilde{a}'Y'\tilde{u} = n(1-R^2) = n\tilde{\sigma}^2$, from which follows that $\sum_{j=1}^{J}\tilde{a}_j \operatorname{cov}\left(y_{\cdot j}, \tilde{u}\right) = (1/n)\tilde{a}'Y'\tilde{u} = \tilde{\sigma}^2$. Thus, the variance of the residual is apportioned among all of the dependent variables, with weights determined by their coefficients.

2.3.2 Second order conditions

Let $\mathcal{L}_2 = \dfrac{\partial^2 \mathcal{L}}{\partial\theta\,\partial\theta'}(\tilde{\theta},\tilde{\mu})$ be the Hessian of the Lagrangian with respect to the parameters, evaluated at the QMLE, and define $g(\theta) = a'Y'Ya - n$ so that that the unit variance constraint may be expressed as $g(\theta) = 0$. A sufficient second order condition for $\mathcal{L}(\tilde{\theta},\tilde{\mu})$ to be a maximum subject to $g(\tilde{\theta}) = 0$ is that $\theta'\mathcal{L}_2\theta < 0$ for $\theta \neq 0$ in the tangent plane $T = \left\{\theta : \nabla g(\tilde{\theta})'\theta = 0\right\}$.[7]

Proposition 3. The QMLE $(\tilde{\theta},\tilde{\mu})$ satisfies the sufficient second order condition for maximizing the constrained likelihood function.

[7] See, e.g., Luenberger (1965, Section 10.3).

2.3.3 Asymptotic distribution of QMLE

To derive the asymptotic distribution of the QMLE, we need to state a series of assumptions with regard to the asymptotic properties of the likelihood and constraint functions, evaluated at the QMLE.

Assumption Set 2:

(i) $\frac{1}{n}E\frac{\partial^2 \log L}{\partial\theta\partial\theta'}(\tilde{\theta})$ converges in probability to a finite nonsingular matrix for any sequence $\left\{\theta^*_{(n)}\right\}$ such that $p\lim\theta^*_{(n)} = \theta$.

(ii) $\frac{1}{n}Y'Y\tilde{a}$ converges in probability to a finite vector.

(iii) $\frac{1}{\sqrt{n}}\frac{\partial \log L}{\partial\theta}(\tilde{\theta})$ converges in probability to $N(0,V)$, where

$$V = \lim E\frac{1}{n}\left[\left(\frac{\partial \log L}{\partial\theta}(\tilde{\theta})\right)\left(\frac{\partial \log L}{\partial\theta}(\tilde{\theta})\right)'\right].$$

Now consider the probability limits of the expected negative Hessian $H_n = \frac{1}{n}E\left[-\frac{\partial^2 \log L}{\partial\theta\,\partial\theta'}(\tilde{\theta})\right]$, the gradient $G_n = 1/n\nabla g(\tilde{\theta})$ of the constraint function $g(\theta) = a'Y'Ya - n$, and the information matrix $\mathcal{I}_n = \frac{1}{n}E\left[\left(\frac{\partial \log L}{\partial\theta}(\tilde{\theta})\right)\left(\frac{\partial \log L}{\partial\theta}(\tilde{\theta})\right)'\right]$, all evaluated at the QMLE. Under some circumstances, for instance when the maximum likelihood specification is correct, the matrices H_n and $\mathcal{I}_n$ are asymptotically the same.[8]

[8] See, e.g., White (1996, Chapter 6).

Proposition 4. Under Assumption Sets 1 and 2, the constrained QMLE $\tilde{\theta}$ is asymptotically normally distributed with $\sqrt{n}\left(\tilde{\theta}-\theta\right) \sim N\left(0, \mathcal{H}_n^{-1}\mathcal{I}_n\mathcal{H}_n^{-1}\right)$, where $\mathcal{H}_n^{-1} = H_n^{-1} - H_n^{-1}G_n\left(G_n' H_n^{-1} G_n\right)^{-1} G_n' H_n^{-1}$. If $H_n = \mathcal{I}_n$, then $\mathcal{H}_n^{-1}\mathcal{I}_n\mathcal{H}_n^{-1} = \mathcal{H}_n^{-1}$.

Using the explicit form of $\log L$ in generalized canonical regression, from equation (4), we may construct the estimates

$$\tilde{H}_n = \frac{1}{n\tilde{\sigma}^2}\begin{bmatrix} Y'Y & -Y'X & -1\big/\tilde{\sigma}^2\, Y'Y\tilde{a} \\ -X'Y & X'X & 0 \\ -1\big/\tilde{\sigma}^2\, \tilde{a}'Y'Y & 0 & n\big/\left(2\tilde{\sigma}^2\right) \end{bmatrix}, \tag{14}$$

$$\tilde{G}_n = 1/n\left[2\tilde{a}'Y'Y \quad 0 \quad 0\right]', \text{ and} \tag{15}$$

$$\tilde{\mathcal{H}}_n^{-1} = \tilde{H}_n^{-1} - \tilde{H}_n^{-1}\tilde{G}_n\left(\tilde{G}_n'\tilde{H}_n^{-1}\tilde{G}_n\right)^{-1}\tilde{G}_n'\tilde{H}_n^{-1}. \tag{16}$$

If the information matrix equality holds, the estimated covariance matrix of $\tilde{\theta}$ is $\frac{1}{n}\mathcal{H}_n^{-1}$. If not, $\mathcal{I}_n$ may be estimated by

$$\tilde{\mathcal{I}}_n = \frac{1}{n\tilde{\sigma}^4}\begin{bmatrix} \sum_{l=-L}^{L} w_l \sum_t y_t\tilde{u}_t\tilde{u}_{t-l}y_{t-l}' & -\sum_{l=-L}^{L} w_l \sum_t y_t\tilde{u}_t\tilde{u}_{t-l}x_{t-l}' & 0 \\ -\sum_{l=-L}^{L} w_l \sum_t x_t\tilde{u}_t\tilde{u}_{t-l}y_{t-l}' & \sum_{l=-L}^{L} w_l \sum_t x_t\tilde{u}_t\tilde{u}_{t-l}x_{t-l}' & 0 \\ 0 & 0 & n/2 \end{bmatrix}, \tag{17}$$

which accommodates various methods proposed, for instance, in White (1980), Hansen (1982), Newey and West (1987) and Andrews (1991) for constructing heteroskedasticity and autocorrelation consistent variances and covariances. The weights w_t are determined by the particular method applied to a lag window of $\pm L$ periods. The estimated covariance matrix of $\tilde{\theta}$ is then $\frac{1}{n}\tilde{\mathcal{H}}_n^{-1}\tilde{\mathcal{I}}_n\tilde{\mathcal{H}}_n^{-1}$.

2.3.4 Variables with non-zero means

If we allow the means of (x_t, y_t) to be non-zero, there are three alternative approaches to generalized canonical regression: demeaning, including a constant term, and suppressing a constant term with non-zero means.

Traditional canonical correlation analysis calls for extracting the sample means of the variables, since the focus is on centered correlations. Demeaning is also an option in generalized canonical correlation. Let $\tilde{y}_t = y_t - \overline{y}$ and $\tilde{x}_t = x_t - \overline{x}$ be the demeaned variables and consider the model

$$\tilde{y}_t{}'a = \tilde{x}_t{}'b + u_t\,. \tag{18}$$

Since this equation has the same properties as (1), it clearly may be estimated using the QMLE techniques presented so far. Written in terms of the original variables, the model becomes

$$y_t{}'a = x_t{}'b + \overline{y}'a - \overline{x}'b + u_t\,, \tag{19}$$

which is the same as (1) except for the addition of an implicit constant term $\overline{y}'a - \overline{x}'b$. If this constant term is acceptable but of no intrinsic interest, demeaning is a reasonable solution.

If there is interest in the value of the constant term itself, or in testing whether it is statistically significantly different from zero in the sample, an explicit constant term may be added to the generalized canonical regression specification. As in a standard regression, adding a constant term has effects similar to demeaning, but allows for calculation and inference with regard to the constant.

Say $x_{\cdot 1}$, the first variable in x_t, is a vector of ones. The orthogonality conditions in the second block of (9) imply that

$$x_{\cdot 1}'u = x_{\cdot 1}'(Ya - Xb) = \sum_{t=1}^{n}\left(y_t'a - x_t'b\right) = 0\,. \tag{20}$$

Thus

$$\overline{y}'a = \frac{1}{n}\sum_{t=1}^{n} y_t'a = \frac{1}{n}\sum_{t=1}^{n} x_t'b = \overline{x}'b\,. \tag{21}$$

Subtracting (21) from each row of (3), the model may be written and estimated in terms of the demeaned values of x_t and y_t, excepting the constant. We may then insert the original (non-demeaned) values of x_t and y_t in equation (11) to obtain the value of the constant, and in (14) to (17) to estimate the asymptotic distribution of the coefficients. Since the constraint is expressed in terms of the variance, note that the term $Y'Y\tilde{a}$ that appears twice in $\tilde{H}_n$ and once in $\tilde{G}_n$ must be replaced by $\left(Y - \overline{Y}\right)'\left(Y - \overline{Y}\right)\tilde{a}$, where $\overline{Y}$ is like Y but with $\overline{y}'$ replacing y_t' in every row.

Demeaning and the introduction of a constant term may be thought of as alternative ways of imposing the assumption that $E(u_t \mid x_t) = 0$ in the empirical model. If there are theoretical reasons in the underlying model that suggest that this assumption should not be imposed, the QMLE setting may still be used with non-demeaned variables and no constant term. However, the analytical expressions developed above are not directly applicable and the correspondence with canonical correlation does not hold. The most practical approach in this case is to solve the likelihood maximization problem numerically.

3. Dynamic generalized canonical regression

A natural application of generalized canonical regression arises in the special case in which the jointly dependent variables in the vector y_t represent leads and lags of a single scalar dependent variable. For example, let P_t be the monthly level of the consumer price index. A variable like $\pi_t^{(12)} = \log(P_{t+12} / P_t)$, average CPI inflation over the year from month *t* to month *t*+12, might be used as the dependent variable in a standard linear regression used for prediction. However, this variable may be expressed as a weighted sum of monthly inflation rates $\pi_t^{(12)} = \sum_{j=1}^{12} a_j \pi_{t+j}^{(1)}$, with weights assumed to be $a_j = 1$ for $j = 1,\ldots,12$. Are the 12 monthly rates equally important in a particular predictive equation, or do weights differ optimally across horizons?

It may be, for instance, that the explanatory variables x_t are best suited for predictions up to 6 months only. One alternative then would be to run the equation with $\pi_t^{(6)}$ as the dependent variable. If the precise predictive lead time is unknown, one could try in turn all possible horizons up to, say, 24 or 36 months.[9]

Generalized canonical regression presents the option of testing all predictive leads simultaneously. In the foregoing example, the jointly dependent variables may be defined as $y_t' = \left(\pi_{t+1}^{(1)},\ldots,\pi_{t+36}^{(1)}\right)$ and dynamic generalized canonical regression may be used to estimate optimal weights a_j for $j = 1,\ldots,36$. The importance of each of the dependent variables can then be assessed by testing the statistical significance of the associated coefficients.

---

[9] The technique of testing multiple individual lead times has been applied, for instance, in Bernard and Gerlach (1997) and Estrella and Mishkin (1997).

More generally, for scalar dependent variables $z_{1t}, z_{2t}, \ldots$, dynamic generalized canonical regression takes the form

$$\left(z_{1t+j_{11}}, z_{1t+j_{12}}, \ldots, z_{2t+j_{21}}, z_{2t+j_{22}}, \ldots\right)' a = x_t{}' b + u_t\,. \tag{22}$$

As the equation suggests, it is possible to have leads and lags of more than one scalar dependent variable and to include nonconsecutive lags. It is also possible to have lags of several regressors on the right hand side.

The format of dynamic generalized canonical regression may raise concerns that the disturbances u_t are serially correlated. As noted earlier, heteroskedasticity and autocorrelation consistent (HAC) estimates of the information matrix $\mathcal{I}_n$ may be computed, for instance, using the methods of Hansen (1982), Newey and West (1987), or Andrews (1991).

4. Empirical illustrations

This section provides three illustrative applications of generalized canonical regression, each of which is analogous to estimates found in the earlier literature. In each case, the dependent variable in the earlier research is the sum of two or more components that receive equal weights. As suggested above, generalized canonical regression is used to allow the weights to differ across components.

4.1 Cross section (panel) application: bank capital ratios

Using annual data for 24 of the 25 largest banking institutions in the United States in 1997, Hirtle (1998) finds that most of them reduced their regulatory capital ratios over the year. One explanation is that they started off the year with high capital ratios, and that there is a tendency for banks to revert to some mean level of the ratio. If this explanation holds, banks with higher ratios should have experienced larger declines in the ratio.

For a given bank, let A_t represent risk-weighted assets and K_t be total capital at the end of year *t*. Hirtle (1998) reports on the regression

$$\Delta r = c_0 + c_1 r_{-1} + \varepsilon , \qquad (23)$$

where $\Delta r = K_t / A_t - K_{t-1} / A_{t-1}$ and $r_{-1} = K_{t-1} / A_{t-1}$ for *t*=1997. The paper finds that the coefficient c_1 is indeed significantly negative, as the proposed explanation would entail.

To obtain a larger sample, the first row of Table 1 shows estimates of equation (23) using annual data for the same 24 banks from 1995 to 1997.[10] The table shows unadjusted standard errors (first entry in parentheses below each coefficient estimate) as well as standard errors allowing for heteroskedasticity as in White (1980) (second entry). The relationship between the dependent variable and the regressor is shown graphically in Figure 1.

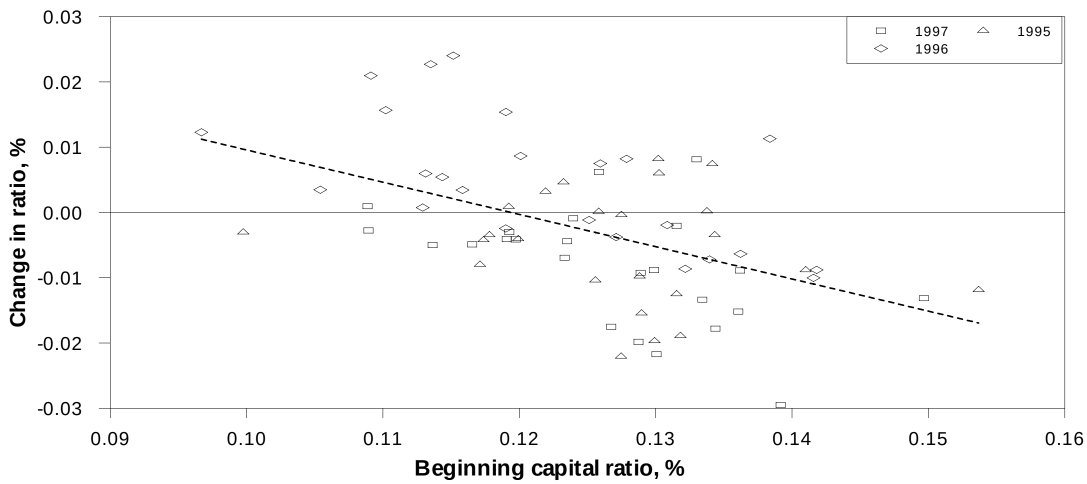
Table 1. Results for bank capital models, 1995-97

Equation	$\Delta\kappa$	$\Delta\kappa_1$	$\Delta\kappa_2$	$\Delta\alpha$	r_{-1}	c_0	R^2
(23)	–	–	–	–	-.49 (0.10,0.10)	.059 (.013,.012)	.243
(25)	88.0 (5.03,6.79)	–	–	-100.8 (4.69,6.33)	-46.8 (9.61,9.13)	5.44 (1.21,1.11)	.247
(27)	–	67.6 (10.4,12.1)	118.2 (11.8,13.5)	-95.5 (5.74,7.76)	-48.3 (9.50,9.70)	5.73 (1.20,1.19)	.264

Note: Standard errors are shown in parentheses. The first is computed under the iid assumption, the second corrects for heteroskedasticity as in White (1980).

[10] The author is grateful to Beverly Hirtle for making available the data from her paper.

Figure 1



Now, if we define

$$\Delta\kappa = \frac{\Delta K_t}{A_{t-1}} \qquad \Delta\alpha = \frac{K_t}{A_t}\frac{\Delta A_t}{A_{t-1}} \tag{24}$$

as the contributions of the numerator and the denominator, respectively, to the change in the capital ratio, we note that $\Delta r = \Delta\kappa - \Delta\alpha$. The corresponding generalized canonical regression is

$$a_1\Delta\kappa + a_2\Delta\alpha \quad = \quad c_0 + c_1 r_{-1} + \varepsilon \tag{25}$$

and we note that equation (23) imposes the restriction $a_2 = -a_1$ on (25).

Estimates of this generalized canonical regression are provided in the second row of Table 1. The scale of the coefficient estimates is different because of the unit variance assumption imposed on the left hand side of the equation, but the significance level of the regressor is clearly about the same, as is the regression R^2 . A statistical test of $a_2 = -a_1$ using the asymptotic distribution of the QMLE cannot reject that the parameters are of equal magnitude. The *p* value is .189, or .331 correcting for heteroskedasticity. In this case, generalized canonical regression does not produce any information that was not contained in the standard regression.

An additional level of detail is obtained by recognizing that total capital is the sum of Tier 1 and Tier 2 components, $K_t = K_t^{(1)} + K_t^{(2)}$. The most important component of Tier 1 capital is equity, whereas Tier 2 capital includes subordinated debt and loan loss reserves. Are these two components equally important in the regression, as equations (23) and (25) maintain? If we now define

$$\Delta\kappa_1 = \frac{\Delta K_t^{(1)}}{A_{t-1}} \qquad \Delta\kappa_2 = \frac{\Delta K_t^{(2)}}{A_{t-1}}, \tag{26}$$

we have that $\Delta r = \Delta\kappa_1 + \Delta\kappa_2 - \Delta\alpha$, which suggests the generalized canonical regression

$$a_1\Delta\kappa_1 + a_2\Delta\kappa_2 + a_3\Delta\alpha \quad = \quad c_0 + c_1 r_{-1} + \varepsilon\,. \tag{27}$$

Equation (23) imposes the restrictions $a_1 = a_2 = -a_3$ on (27).

Results for this equation appear in the third row of Table 1. Estimates of a_3 and c_1 are not very different from their counterparts in (25), but we see here that the weights on the two components of capital are very different. The test of $a_1 = a_2$ has *p* value of .011. (.016 correcting for heteroskedasticity) and the test of $a_1 = a_2 = -a_3$ has *p* value of .015 (.035 correcting for heteroskedasticity). Thus, generalized canonical regression indicates that a weighted sum of changes in Tier 1 and Tier 2 capital is most predictable when significantly greater weight is assigned to Tier 2 capital. At this level of detail, generalized canonical regression uncovers more information than standard regression.

4.2 Dynamic generalized canonical regression (1): Yield curve and growth

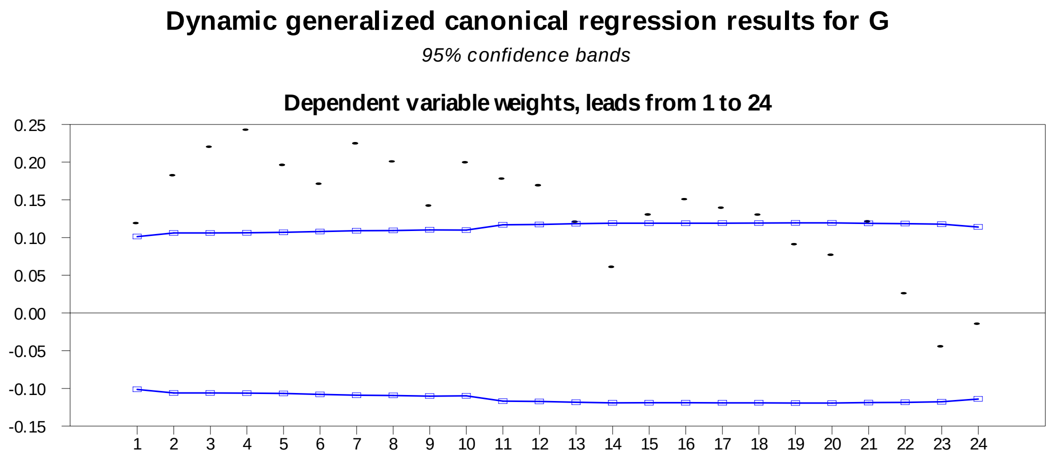
For simplicity, we focus on two examples of dynamic generalized canonical regression in which there is only one scalar dependent series and one regressor, since such examples suffice to illustrate the distinctive properties of generalized canonical regression. Consider first the equation

$$E\left(\sum_{j=1}^{24} a_j y_{t+j}^{(1)} \mid s_t\right) = b_1 s_t + b_2\,, \tag{28}$$

where $y_t^{(1)}$ is growth (first difference of log) in industrial production in month t and s_t is the difference between the 10-year U.S. Treasury constant maturity rate and the 3-month secondary-market Treasury rate on a bond equivalent basis. Rates are monthly averages of daily data. This equation is analogous to the two-year cumulative equation in Estrella, Rodrigues and Schich (2003, Table 4), but here the weights on the monthly growth rates are allowed to differ.

The sample period is from February 1959 to September 2003. The early end of the sample is necessary to accommodate a maximum predictive horizon of 36 months, used in the next example. As in earlier literature using equal weights, the term spread is significant. The parameter b_1 (taken to be positive) is estimated as .361 with a standard error of .030 and the regression R^2 is .216. Rather than list the coefficient estimates of the 24 jointly dependent variables and their standard errors, Figure 2 presents the coefficient values graphically, along with asymptotic 95% confidence intervals.

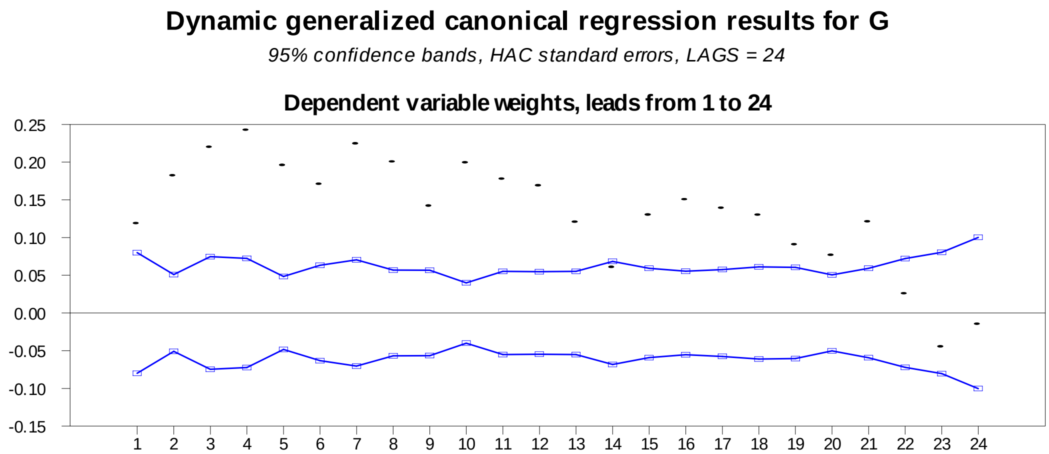
Figure 2



The pattern of the dependent variable weights is roughly consistent with using a 12-month average as the single dependent variable, which is relatively standard in the literature. However, equality of the weights is formally rejected at the 5% level. For instance, the restriction that the a_j are equal has an asymptotic *p* value of .001 for *j* = 1, …, 24 and of .028 for *j* = 1, …, 12. The restriction that $a_j = 0$ for *j* = 13, …, 24 has a *p* value of .000.

As in standard regression, one important concern in this type of predictive equation is that overlapping predictive horizons may give rise to moving-average errors. Thus, we apply a Newey-West (1987) correction with 24 lags. The standard error of b_1 is .102, higher than without the adjustment but still indicative of statistical significance. HAC standard errors for many of the estimated dependent variable coefficients are lower than the unadjusted values, as shown in Figure 3.

Figure 3



With HAC standard errors, the qualitative conclusions about equality of the dependent variable coefficients remain the same. The restriction that the a_j are equal has an asymptotic *p* value of .000 for *j* = 1, …, 24 and of .001 for *j* = 1, …, 12. The restriction that $a_j = 0$ for *j* = 13, …, 24 has a *p* value of .000.

4.3 Dynamic generalized canonical regression (2): Yield curve and inflation

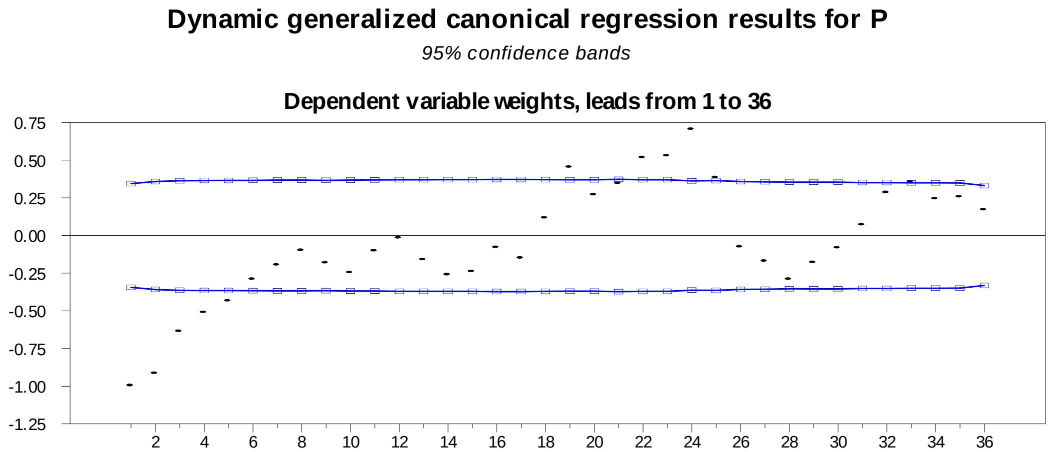
An analogous equation for inflation is

$$E\left(\sum_{j=1}^{36} a_j \pi_{t+j}^{(1)} \mid s_t\right) = b_1 s_t + b_2\,, \tag{29}$$

where $\pi_t^{(1)}$ is CPI inflation (first difference of log CPI) in month *t* and s_t is as before. This equation is similar to estimates in Estrella and Mishkin (1997), but uses monthly instead of quarterly data.

Again, the sample period is from February 1959 to September 2003. As in related earlier research, the term spread is found to be significant in this equation. The parameter b_1 (taken to be positive) is estimated as .435 with a standard error of .028. The regression R^2 is .313. Estimates of the dependent variable coefficients are show in Figure 4.

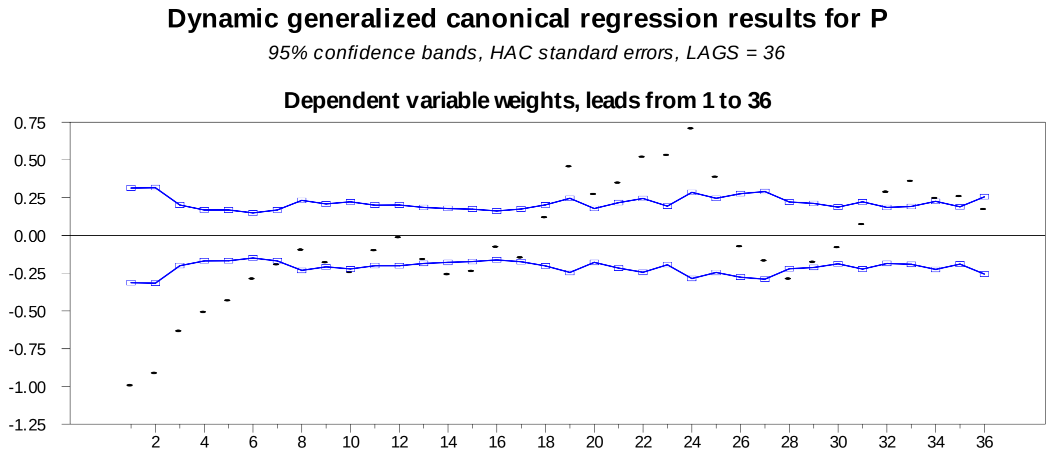
Figure 4



The pattern of weights that emerges for the dependent variables is generally increasing, which is consistent with economic theory. For instance, theoretical models in Mishkin (1990) and Estrella (2005) suggest that the term spread should be a useful predictor of increases in inflation, rather than levels. In both cases, the term spread is seen as predicting changes of the form $(1/n)\pi_{t+m}^{(n)} - (1/m)\pi_t^{(m)}$, which is roughly consistent with Figure 4 where we see that early leads have significantly negative coefficients and leads in the vicinity of two years have significantly positive coefficients. However, a strict hypothesis of this form with equal weights in each term is rejected by the data.

With Newey-West HAC correction with 36 lags, the standard error for b_1 is .135, again higher that the unadjusted estimate but indicative of statistical significance. Standard errors are generally lower for the dependent variables, as shown in Figure 5, though the qualitative conclusions about parameter significance are similar.

Figure 5



5. Conclusions

Generalized canonical regression extends the format of standard regression by allowing the left side of the equation to consist of a linear combination of jointly dependent variables. Natural applications arise when the dependent variable in a standard regression is the sum of equally-weighted terms that could conceivably have unequal optimal weights. To illustrate, the paper provides examples using cross-sectional or panel data, as well as examples of dynamic generalized canonical regressions in which the dependent variables are leads of a single series.

In contrast to the earlier literature, this paper constructs an estimate of the asymptotic distribution of the QMLE estimates, which allows for statistical inference with regard to the parameters. Moreover, the asymptotic distribution of the parameters is consistent in the presence of heteroskedasticity and autocorrelation.

Appendix: Proofs

Proposition 1. The proof is analogous to the closely related Corollary 5.3 in White (1996). The main difference is that we consider the conditional distribution of a linear combination of the dependent variables, rather than of all the individual dependent variables (Assumptions 1(ii) and 1(iii)). Moreover, the likelihood function is maximized with regard to the linear combination of the dependent variables, as well as with regard to the linear combination of regressors (Assumption 1(vii)).

Proposition 2. From equation (13), we see that $1+2\tilde{\mu}\tilde{\sigma}^2$ is an eigenvalue of $\left(Y'Y\right)^{-1}Y'X\left(X'X\right)^{-1}X'Y$ or, equivalently, of $\left(Y'Y\right)^{-1/2}Y'X\left(X'X\right)^{-1}X'Y\left(Y'Y\right)^{-1/2}$. The latter matrix is symmetric and positive definite, hence all its eigenvalues are real and positive. Premultiplying (12) by a' and dividing by $\tilde{a}'Y'Y\tilde{a}$ (= *n*),

$$1+2\tilde{\mu}\tilde{\sigma}^2=\tilde{a}'Y'X\left(X'X\right)^{-1}X'Y\tilde{a}\left(\tilde{a}'Y'Y\tilde{a}\right)^{-1}=\tilde{b}'X'X\tilde{b}\left(\tilde{a}'Y'Y\tilde{a}\right)^{-1}=R^2, \quad (30)$$

where the last equality defines R^2 as the proportion of the variance of the linear combination in the left hand side of the equation that is explained by the linear combination in the right hand side. Thus, we take $1+2\tilde{\mu}\tilde{\sigma}^2=R^2=r_1$, where r_1 is the largest eigenvalue of $\left(Y'Y\right)^{-1}Y'X\left(X'X\right)^{-1}X'Y$. The estimate $\tilde{a}$ is the eigenvector corresponding to the largest eigenvalue, scaled to satisfy constraint (6).

To complete the solution of the first order conditions, we compute the variance of the disturbance and the Lagrange multiplier. Equations (6) and (30) imply that

$$\tilde{b}'X'X\tilde{b} = \tilde{a}'Y'X\tilde{b} = nR^2\,. \tag{31}$$

Hence, $\tilde{u}'\tilde{u} = n(1-R^2)$, from which it follows that $\tilde{\sigma}^2 = 1-R^2$ and $\tilde{\mu} = -1/2$.

Proposition 3. In the case of generalized canonical regression,

$$\mathcal{L}_2 = \begin{bmatrix} -\tilde{\sigma}^{-2}R^2Y'Y & \tilde{\sigma}^{-2}Y'X & \tilde{\sigma}^{-4}Y'Y\tilde{a} \\ \tilde{\sigma}^{-2}X'Y & -\tilde{\sigma}^{-2}X'X & 0 \\ \tilde{\sigma}^{-4}\tilde{a}'Y'Y & 0 & -\tilde{\sigma}^{-4}n/2 \end{bmatrix} \tag{32}$$

and $T = \left\{\theta : \nabla g(\tilde{\theta})'\theta = 0\right\} = \left\{a : \tilde{a}'Y'Ya = 0\right\}$. Thus, we need that

$$\begin{aligned} \theta'\mathcal{L}_2\theta &= -\tilde{\sigma}^{-2}\left(R^2a'Y'Ya - 2a'Y'Xb + b'X'Xb\right) - \tilde{\sigma}^{-4}\sigma^4 n/2 \\ &= -\tilde{\sigma}^{-2}\left(u'u - (1-R^2)a'Y'Ya\right) - \tilde{\sigma}^{-4}\sigma^4 n/2 \end{aligned} \tag{33}$$

is negative for nonzero $\theta \in T$. Since $1-R^2 = \tilde{u}'\tilde{u}(\tilde{a}'Y'Y\tilde{a})^{-1}$ and the QMLE minimizes $u'u(a'Y'Ya)^{-1}$, we have that $u'u - (1-R^2)a'Y'Ya \geq 0$. Equality holds only if $a = \lambda\tilde{a}$ for real $\lambda \neq 0$, but then $a \notin T$. Hence, the strict inequality holds for all $a \in T$, as required by the second order condition.

Proposition 4. The proof follows from Aitchison and Silvey (1958), with one modification. Aitchison and Silvey (1958) assume from the outset that the information matrix equality $H_n = \mathcal{I}_n$ holds and show that the asymptotic variance of $\sqrt{n}\tilde{\theta}$ is $\mathcal{H}_n^{-1}$. Refraining from this assumption and following the same steps as in that paper leads to the more general result $\mathcal{H}_n^{-1}\mathcal{I}_n\mathcal{H}_n^{-1}$. Gallant (1987, Section 3.7) has a similar result.

References

Aitchison, J. and Silvey, S.D. (1958), "Maximum-likelihood estimation of parameters subject to restraints," Annals of Mathematical Statistics, 29, 813-828.

Anderson, T.W. (1999), "Asymptotic theory for canonical correlation analysis," Journal of Multivariate Analysis, 70, 1-29.

Anderson, T.W. (2002) "Canonical correlation analysis and reduced rank regression in autoregressive models," Annals of Statistics, 30, 1134-1154.

Andrews, Donald W.K. (1991), "Heterokedasticity and autocorrelation consistent covariance matrix estimation," Econometrica, 59, 817-858.

Bartlett, M.S. (1938), "Further aspects of the theory of multiple correlation," Proceedings of the Cambridge Philosophical Society, 34, 33-40.

Bartlett, M.S. (1951), "The effect of standardization on a χ^2 approximation in factor analysis," Biometrika, 38, 337-344.

Bernard, Henri and Stefan Gerlach (1997), "Does the term structure predict recessions? The international evidence," International Journal of Finance and Economics, 3, 195-215.

Box, G.E.P. and Jenkins, G.M. (1970), Time Series Analysis, Forecasting and Control, San Francisco: Holden Day.

Breiman, Leo and Freidman, Jerome H. (1985), "Estimating optimal transformations for multiple regression and correlation," Journal of the American Statistical Association, 80, 580-598.

Engle, Robert F. (1980), "Exact maximum likelihood methods for dynamic regressions and band spectrum regressions," International Economic Review, 21, 391-407.

Estrella, Arturo (2005), "Why does the yield curve predict output and inflation?" Economic Journal, 115, 722-744.

Estrella, Arturo and Mishkin, Frederic S. (1997), "The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank," European Economic Review, 41, 1375-1401.

Estrella, Arturo, Rodrigues, Anthony P. and Schich, Sebastian (2003), "How stable is the predictive power of the yield curve? Evidence from Germany and the United States," Review of Economics and Statistics, 85, 629-644.

Fisher, Franklin M. (1966), The Identification Problem in Econometrics, Huntington, New York: Robert E. Krieger Publishing Co.

Gallant, A. Ronald (1987), Nonlinear Statistical Models, New York: John Wiley & Sons.

Gyimah-Brempong, Kwabena and Gyapong, Anthony O. (1991), "Characteristics of education production functions: An application of canonical regression analysis," Economics of Education Review, 10, 7-17.

Hansen, Lars Peter (1982), "Large sample properties of generalized method of moments estimators," Econometrica, 50, 1029-1054.

Hirtle, Beverly (1998), "Bank holding company capital ratios and shareholder payouts," Federal Reserve Bank of New York Current Issues in Economics and Finance, 4(9), 1-6.

Hotelling, Harold (1935), "The most predictable criterion," Journal of Educational Psychology, 26, 139-142.

Hotelling, Harold (1936), "Relations between two sets of variates," Biometrika, 28, 321-377.

Luenberger, David G. (1965), Introduction to Linear and Nonlinear Programming, Reading, MA: Addison Wesley.

Margulis, Harry L. (1998), "Predicting the growth and filtering of at-risk housing: Structure, ageing, poverty and redlining," Urban Studies, 35, 1231-1259.

Mishkin, Frederic S. (1990), "What does the term structure tell us about future inflation?" Journal of Monetary Economics, 25, 77-95.

Newey, W.K. and West, K.D. (1987), "A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," Econometrica, 55, 703-708.

Ruggiero, John (1996), "Efficiency of educational production: An analysis of New York school districts," The Review of Economics and Statistics, 78, 499-509.

Theil, Henri (1971), Principles of Econometrics, New York: John Wiley & Sons.

Tintner, Gerhard (1946), "Some applications of multivariate analysis to economic data," Journal of the American Statistical Association, 41, 472-500.

Tintner, Gerhard (1950), "Some formal relations in multivariate analysis," Journal of the Royal Statistical Society, Series B (Methodological), 12, 95-101.

Trippi, Robert R. (1977), "Estimating the relationship between price and time to sale for investment property," Management Science, 23, 838-842.

Vinod, H.D. (1969), "Econometrics of joint production—A reply," Econometrica, 37, 739-740.

Waugh, Frederick V. (1942), "Regressions between sets of variables," Econometrica, 10, 290-310.

White, Halbert (1980), "A heteroskedasticity consistent covariance matrix estimator and a direct test of heterskedasticity," Econometrica, 48, 817-838.

White, Halbert (1996), Estimation, Inference and Specification Analysis, Cambridge: Cambridge University Press.