Department of Economics
Working Paper Series

Projection Minimum Distance: An Estimator for Dynamic Macroeconomic Models

Oscar Jorda
University of California, Davis

Oscar Jorda
University of California, Davis

Sharon Kozicki
Bank of Canada

August 22, 2006

Paper # 06-27

This paper introduces an estimator for dynamic macroeconomic models where possibly the dynamics and the variables described therein are incomplete representations of a larger, unknown macroeconomic system. We call this estimator projection minimum distance (PMD) and show that it is consistent and asymptotically normal. Many times, PMD can provide consistent estimates of structural parameters even when the dynamics of the macroeconomic model are insufficient to account for the serial correlation of the data or correlation with information omitted from the model. PMD provides an overall specification chi-squared test based on the distance between the impulse responses of the model and their semi-parametric estimates from the data. PMD only requires two, simple, least-squares steps and can be generalized to more complex, nonlinear environments.

UCDAVIS

Department of Economics
One Shields Avenue
Davis, CA 95616
(530)752-0741

http://www.econ.ucdavis.edu/working_search.cfm

This version, August 2006

Projection Minimum Distance: An Estimator for Dynamic Macroeconomic Models*

Abstract

This paper introduces an estimator for dynamic macroeconomic models where possibly the dynamics and the variables described therein are incomplete representations of a larger, unknown macroeconomic system. We call this estimator *projection minimum distance* (PMD) and show that it is consistent and asymptotically normal. Many times, PMD can provide consistent estimates of structural parameters even when the dynamics of the macroeconomic model are insufficient to account for the serial correlation of the data or correlation with information omitted from the model. PMD provides an overall specification chi-squared test based on the distance between the impulse responses of the model and their semi-parametric estimates from the data. PMD only requires two, simple, least-squares steps and can be generalized to more complex, nonlinear environments.

- *Keywords:* impulse response, local projection, rational expectations, minimum chi-square, minimum distance.
- *JEL Codes:* C32, E47, C53.

Òscar Jordà
Department of Economics
University of California, Davis
One Shields Ave.
Davis, CA 95616
e-mail: ojorda@ucdavis.edu

Sharon Kozicki
Research Department
Bank of Canada
234 Wellington Street
Ottawa, Ontario, Canada
K1A 0G9
e-mail: skozicki@bank-banque-canada.ca

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Bank of Canada. We thank Colin Cameron, Timothy Cogley, David Drukker, Jeffrey Fuhrer, James Hamilton, Peter Hansen, Kevin Hoover, Giovanni Olivei, Frank Wolak and seminar participants at Duke University, the Federal Reserve Bank of New York, Federal Reserve Bank of San Francisco, Federal Reserve Bank of St. Louis, Stanford University, the University of California, Davis, Univeristy of California, Riverside, University of Kansas and the Winter Meetings of the American Economic Association, 2006 in Boston for useful comments and suggestions. Jordà is thankful for the hospitality of the Federal Reserve Bank of San Francisco during the preparation of this paper.

1 Introduction

Estimating and testing alternative structural models of the macroeconomy is necessary to advance our understanding of the fundamental forces at work and it is critical in establishing appropriate policy responses in actual economies. Inevitably, macroeconomic models compromise realism in favor of tractability and analysis. These opposing forces offer significant challenges for formal statistical evaluation. For instance, the maximum likelihood principle and its asymptotic optimality properties require that the structural model be a representation of the density of the underlying data generation process (DGP). This demand is very hard to meet in practice and results in rejection of many economically useful models. Routine failure of specification tests has therefore led researchers down the path of evaluating economic models more informally.

This paper introduces new methods to estimate and evaluate dynamic, stochastic macroeconomic models, which may only be partial representations of a larger, unknown, macroeconomic system. The method, which we label projection minimum distance (PMD), is a two-step estimator. In the first step, we estimate impulse responses from the data semiparametrically with the local projections estimator introduced by Jordà (2005). These impulse responses can be calculated from a system with many variables that are not included nor explained by the candidate macroeconomic model. Next, we represent the stable solution of the model in terms of the Wold representation of this larger system, and obtain the mapping between the structural parameters and the Wold coefficients by the method of undetermined coefficients (see Christiano, 2002). Hence, the structural parameters of the model are estimated in a second step that consists of minimizing the distance between the impulse responses from the data and those implied by the model. The resulting estimator is based on a minimum chi-square estimator (Ferguson, 1958) and belongs to the broader family of minimum distance estimators of which GMM is also a member of the class.

However, PMD has important advantages that distinguish it from GMM and other commonly used estimators. First, we provide an overall misspecification test based on overidentifying restrictions that is distributed chi-square. Effectively, this test formalizes the common practice of evaluating a model by how well its impulse responses match those from the data. Second, PMD provides consistent estimates even when the model's dynamics are insufficient to fit the data. We show that the common GMM practice of using lags of the endogenous variables as instruments can only be justified by the internal dynamics of the data, not the dynamics prescribed by the model. This basic, well-known observation is often ignored, which results in invalid instrument problems and inconsistent estimates. PMD avoids these problems since the starting premise is that the model is an insufficient representation for the data. On a practical level, PMD consists of two, simple, least-squares steps, and therefore, is easily implementable and can be generalized to nonlinear environments. In fact, we show that PMD can be used to estimate generic VARMA(p,q) models that would usually require numerical optimization routines.

Econometrically, the paper has two main contributions. First, the consistency and asymptotic normality of PMD requires that we determine the asymptotic distribution of the first-step estimates of the impulse responses from the data. In and of itself, this is an important result as it provides analytic formulas for the asymptotic covariance matrix of the impulse response coefficients across time and across variables. This consistency and asymptotic normality proof accommodates a DGP with possibly infinite lags. Second, the minimum chi-square step is based on an unknown function of the structural parameters that can only be estimated consistently. Therefore, we derive the consistency and asymptotic normality of the minimum chi-square step so that the asymptotic covariance matrix reflects this estimation uncertainty. In addition, we show that an overall misspecification test based on overidentifying restrictions is distributed chi-square.

We introduce PMD and the main results in the context of a flexible, linear state-space representation of a dynamic rational expectations model. However, PMD is not limited by linearity: the first-step local projections can be estimated more flexibly (even nonparametrically) as described in Jordà (2005) and the second step is not limited to linearity nor rational expectations mechanisms. However, because the paper is already dense with results, we defer these developments to future papers.

The empirical section of the paper includes a Monte Carlo exercise where a traditional ARMA(1,1) model is estimated simultaneously by maximum likelihood and by PMD. This exercise is meant to highlight that PMD is truly a general estimation method. PMD is less efficient than maximum likelihood by construction (the efficiency bound is reached in the extreme case in which infinite impulse response coefficients are used) but we show that PMD is quite efficient even in small samples. The second empirical exercise replicates the analysis in Fuhrer and Olivei (2004) and includes PMD as an alternative for comparison. We show that estimates of an IS and Phillips curves by PMD compare very favorably with GMM, maximum-likelihood and optimal-instruments GMM.

2 Projection Minimum Distance: The Method

This section describes the basics of PMD with a backward-forward looking type of formulation. We think of this benchmark model as a summary of the Euler conditions implied by a generic, dynamic, stochastic macroeconomic model of interest. This model may be incomplete in two dimensions: the dynamics of the model may insufficiently explain the dynamics observed in the data, and the model may only describe a subset of the many relevant variables in a macroeconomy. We present the method with this benchmark specification as a scenario empirical researchers are likely to encounter in practice. However, PMD is quite general and we expect that the reader will be able to extrapolate the principles we are about to present to other problems that we do not directly discuss here.

The principle behind PMD consists in representing the stable solution path of the candidate macroeconomic model in terms of its Wold decomposition and the structural parameters we want to estimate. Then we minimize the weighted quadratic distance between the data's and the model's Wold coefficients by choosing the parameter vector that achieves the minimum of this distance.

Representing the solution path in terms of the Wold decomposition is advantageous for two reasons: we do not have to make choices about the roots of the autoregressive representation of the stable path and, in a linear model, the relation between the Wold coefficients and the parameters is linear and uniquely determined. In what follows, we provide a mathematical characterization of this principle that allows us to derive the statistical results that follow.

Consider an economy that can be described by a vector of variables $\mathbf{y}_t = (\mathbf{y}'_{1t} \quad \mathbf{y}'_{2t})'$ of dimension $r \times 1$, where $r = r_1 + r_2$. The proposed macroeconomic model describes the behavior of, possibly, only some of the variables in this system. Without loss of generality, we find it useful collect these variables into the vector $\mathbf{y}_{1t}$ and collect the variables not described by the model into $\mathbf{y}_{2t}$. A natural benchmark is to characterize the behavior of $\mathbf{y}_{1t}$ by a generic rational expectations formulation (e.g., see Farmer, 1993; and Evans and Honkapohja, 2001) with backward- and forward-looking terms, specifically

$$\Phi'_0 \mathbf{y}_{1t} = \Phi'_1 \mathbf{y}_{1t-1} + \Phi'_2 E_t \mathbf{y}_{1t+1} + \mathbf{u}_{1t}, \quad E(\mathbf{u}_{1t}\mathbf{u}'_{1t}) = I \tag{1}$$

where $\mathbf{u}_{1t}$ is the $r_1 \times 1$ vector of expectational errors. The $r_1 \times r_1$ coefficient matrix Φ_0 makes explicit the nature of the contemporaneous relations between elements of $\mathbf{y}_{1t}$. Expression (1) does not imply that the model must have first order dynamics. Nothing in the derivations that follow require that the dynamics be restricted to one lag of $\mathbf{y}_{1t}$: the vector $\mathbf{y}_{1t}$ can always be appropriately redefined so that (1) can be thought of as a state-space representation.[1] When $\mathbf{y}_{1t} = \mathbf{y}_t$, expression (1) completely specifies the economy but otherwise, it should be clear that $\mathbf{y}_{2t}$ collects variables omitted in the model but possibly relevant in the real economy.

A stable solution of the underlying system describing $\mathbf{y}_t$ is a dynamic, stochastic, difference equation. The stability of the solution implies that it is covariance-stationary, and hence, by the Wold decomposition theorem (see Anderson, 1994), can be represented as (for simplicity we omit

[1] We also remark that $\mathbf{y}_{1t}$ can also contain exogenous forcing variables. In that case it should be clear that rows of Φ_2 corresponding to the forcing variables will be zero.

the constant and other deterministic components),

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j' \boldsymbol{\varepsilon}_{t-j}$$

$$\begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix} + \begin{bmatrix} \underset{r_1 \times r_1}{B_{11}^{1'}} & \underset{r_1 \times r_2}{B_{21}^{1'}} \\ \underset{r_2 \times r_1}{B_{12}^{1'}} & \underset{r_2 \times r_2}{B_{22}^{1'}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_{1t-1} \\ \boldsymbol{\varepsilon}_{2t-1} \end{bmatrix} + ... \tag{2}$$

with $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_{1t}' \quad \boldsymbol{\varepsilon}_{2t}')'$, $B_0 = I_r$, $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Sigma_\varepsilon$ where

$$\Sigma_\varepsilon = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and $E(\boldsymbol{\varepsilon}_{it} \boldsymbol{\varepsilon}_{jt-k}) = \mathbf{0}$ for $i \neq j, k \neq 0$. The Wold decomposition for $\mathbf{y}_{1t}$ is therefore

$$\mathbf{y}_{1t} = \boldsymbol{\varepsilon}_{1t} + \sum_{j=1}^{\infty} B_{11}^{j'} \boldsymbol{\varepsilon}_{1t-j} + \sum_{j=1}^{\infty} B_{21}^{j'} \boldsymbol{\varepsilon}_{2t-j}. \tag{3}$$

Substituting this expression into (1) we obtain

$$\begin{aligned} \Phi_0' \left(\boldsymbol{\varepsilon}_{1t} + B_{11}'^{1} \boldsymbol{\varepsilon}_{1t-1} + ... + B_{21}^{1'} \boldsymbol{\varepsilon}_{2t-1} + ... \right) &= \qquad (4) \\ = \Phi_1' \left(\boldsymbol{\varepsilon}_{1t-1} + B_{11}'^{1} \boldsymbol{\varepsilon}_{1t-2} + ... + B_{21}^{1'} \boldsymbol{\varepsilon}_{2t-2} + ... \right) &+ \\ + \Phi_2' \left(B_{11}'^{1} \boldsymbol{\varepsilon}_{1t} + ... + B_{21}^{1'} \boldsymbol{\varepsilon}_{2t} + ... \right) &+ \mathbf{u}_{1t}. \end{aligned}$$

This expression maps the structural coefficients of the macroeconomic model to the impulse response coefficients of the system $\mathbf{y}_t$.

Our next objective is to write down this mapping more specifically, beginning with the coefficients in Φ_0. Hence, consider post-multiplying both sides of expression (4) by $\boldsymbol{\varepsilon}_{1t}'$ and then take expectations on both sides to obtain,

$$\Phi_0' \Sigma_{11} = \Phi_2' B_{11}^{1'} \Sigma_{11} + E(\mathbf{u}_{1t} \boldsymbol{\varepsilon}_{1t}').$$

Let $P_1\mathbf{u}_{1t} = \boldsymbol{\varepsilon}_{1t}$, that is, the reduced-form residuals are simply some rotation of the structural residuals, and noting that Σ_{11}^{-1} is guaranteed to exist and $\Sigma_{11} = P_1P_1'$, the previous expression can be rearranged as

$$\begin{aligned} P_1'\left(P_1P_1'\right)^{-1} &= \Phi_0' I_{r_1} + \Phi_1'\mathbf{0} - \Phi_2' B_{11}^{1'} \\ P_1^{-1} &= \Phi_0' I_{r_1} + \Phi_1'\mathbf{0} - \Phi_2' B_{11}^{1'}. \end{aligned} \tag{5}$$

We now set these conditions aside momentarily to make our derivations more transparent to the reader. In practice, we have found many models can be estimated by ignoring expression (5) with little loss in efficiency. In addition and to further streamline the presentation, we will assume in what follows that $\Phi_0 = I$, as is commonly done in many popular macroeconomic specifications. However, once we establish the basic results with these restrictions, we will derive the results in full generality in section 4.3.

With these considerations, post-multiply expression (4) by $\boldsymbol{\varepsilon}_{1t-j}'$ and $\boldsymbol{\varepsilon}_{2t-j}'$ and then take expectations to arrive at the set of conditions,

$$\left.\begin{aligned} B_{11}^{j'} &= \Phi_1' B_{11}^{j-1'} + \Phi_2' B_{11}^{j+1'} \\ B_{21}^{j'} &= \Phi_1' B_{21}^{j-1'} + \Phi_2' B_{21}^{j+1'} \end{aligned}\right\} \text{ for } j \quad \geq \quad 1 \tag{6}$$

with $B_{11}^0 = I_{r_1}$ and $B_{21}^0 = \mathbf{0}_{r_2,r_1}$. In what follows, the notation $\mathbf{0}_{j,k}$ is used to indicate a matrix of zeros of dimension $j \times k$. The conditions in expression (6) can be stacked conveniently. For that purpose, let

$$B_j = \begin{bmatrix} B_{11}^j & B_{12}^j \\ B_{21}^j & B_{22}^j \end{bmatrix}; \quad B(0,h) = \mathbf{B} = \begin{bmatrix} I_r \\ B_1 \\ \vdots \\ B_h \end{bmatrix}, \tag{7}$$

where h is the impulse responses' maximum horizon considered in the estimation. Notice that h has to be finite in finite samples but can be let to grow as the sample size grows to infinity Define the selector matrices

$$\begin{aligned} S_0 &= \begin{bmatrix} \underset{r(h-1)\times r}{\mathbf{0}} & (I_{h-1}\otimes I_r) & \underset{r(h-1)\times r}{\mathbf{0}} \end{bmatrix}, \\ S_1 &= \begin{bmatrix} (I_{h-1}\otimes I_r) & \underset{r(h-1)\times r}{\mathbf{0}} & \underset{r(h-1)\times r}{\mathbf{0}} \end{bmatrix}, \\ S_2 &= \begin{bmatrix} \underset{r(h-1)\times r}{\mathbf{0}} & \underset{r(h-1)\times r}{\mathbf{0}} & (I_{h-1}\otimes I_r) \end{bmatrix}, \end{aligned} \tag{8}$$

and

$$R = \begin{bmatrix} I_{r_1} \\ \mathbf{0}_{r_2,r_1} \end{bmatrix}. \tag{9}$$

These auxiliary matrices allow us to write the conditions in (6) for $j = 1, ..., h$ more compactly as

$$S_0\mathbf{B}R = S_1\mathbf{B}R\Phi_1 + S_2\mathbf{B}R\Phi_2. \tag{10}$$

The vec operator can be applied to both sides of this expression to appropriately vectorize the coefficient vectors. Let $b \equiv vec(B)$; $\phi \equiv \{vec(\Phi_1) \quad vec(\Phi_2)\}'$ and noting the following relationships,

$$\begin{aligned} vec(S_0\mathbf{B}R) &= (R'\otimes S_0)\, b \qquad (11) \\ vec\left[(S_1\mathbf{B}R \quad S_2\mathbf{B}R)\begin{pmatrix}\Phi_1 \\ \Phi_2\end{pmatrix}\right] &= (I_{r_1}\otimes (S_1\mathbf{B}R \quad S_2\mathbf{B}R))\,\phi \end{aligned}$$

then (10) can be written as

$$(R'\otimes S_0)\, b = \{(I_{r_1}\otimes S_1\mathbf{B}R) \quad (I_{r_1}\otimes S_2\mathbf{B}R)\}\,\phi\,. \tag{12}$$

or in the special case where the system is completely specified and $R = I_r$,

$$(I_r\otimes S_0)\, b = \{(I_r\otimes S_1\mathbf{B}) \quad (I_r\otimes S_2\mathbf{B})\}\,\phi\,. \tag{13}$$

If one had estimates $\widehat{\mathbf{B}}_T$ (and therefore $\widehat{b}_T$) of $\mathbf{B}$ (and therefore b), then expression (12) is of the form $\widehat{b}_T = g\left(\widehat{b}_T;\phi\right)$ and classical minimum-distance estimation would be a natural way to obtain $\widehat{\phi}_T$. In particular, let

$$\begin{aligned} S\widehat{b}_T &\equiv (R' \otimes S_0)\,\widehat{b}_T & (14)\\ g\left(\widehat{b}_T;\phi\right) &\equiv \left\{\left(I_{r_1}\otimes S_1\widehat{\mathbf{B}}_T R\right) \quad \left(I_{r_1}\otimes S_2\widehat{\mathbf{B}}_T R\right)\right\}\phi & (15)\end{aligned}$$

then ϕ can be found by minimizing

$$\min_{\phi} \widehat{Q}_T(\phi) = \left[S\widehat{b}_T - g\left(\widehat{b}_T;\phi\right)\right]' \widehat{W}\left[S\widehat{b}_T - g\left(\widehat{b}_T;\phi\right)\right] \quad (16)$$

for some weighting matrix $\widehat{W}$.

Notice that the matrices Φ_1 and Φ_2 contain $2r_1^2$ parameters that we want to estimate but we have $(h-1)(r_1^2 + r_1 r_2)$ conditions available for estimation.

In the next section we derive consistency and asymptotic normality results for the first-stage local projection estimator proposed therein. Estimates from this first-step are then incorporated into the minimum chi-square step (16), whose consistency and asymptotic normality properties we derive in subsequent sections.

3 First-Step: Local Projections

The first step in deriving the minimum distance estimator of expression (16) is to obtain estimates of the Wold coefficients in $\mathbf{B}$. There are several reasons why we find local projections superior to estimates of $\mathbf{B}$ derived from a finite order VAR. As we will show momentarily, local projections ensure the consistency of $\widehat{\mathbf{B}}$ even when the underlying process is of infinite order. This is an important consideration since an essential class of macroeconomic models have solutions characterized by $VARMA(p,q)$ dynamics. In addition Jordà (2005) shows that the local nature of the approximation of the projections in many cases provides estimates of $\mathbf{B}$ robust to misspecification.

When the underlying dynamics are nonlinear, the possibility of estimating local projections with nonlinear and even nonparametric techniques affords a considerable advantage over $VARs$.

An estimate of the full covariance matrix of $\mathbf{B}$ is another essential element to obtain analytic standard errors in the second-stage, minimum chi-square step. Local projections provide a simple analytic expression for this covariance matrix – that is, the covariance of impulse response coefficients across time and across variables. Estimates of impulse responses based on a VAR require delta-method or numerical simulation techniques to compute their covariance matrix: a very substantial and complex computational burden. Thus, this section derives consistency and asymptotic normality results for the projections that we use in deriving the formal statistical properties of PMD in section 4.

3.1 Consistency

We assume the rational expectations model in (1) has a stable solution. Thus, this model is covariance-stationary and has a Wold decomposition,

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j \boldsymbol{\varepsilon}_{t-j} \tag{17}$$

where for simplicity and without loss of generality we drop the constant and any deterministic terms. From the Wold decomposition theorem (see e.g. Anderson, 1994):

(i) $E(\boldsymbol{\varepsilon}_t) = 0$ and $\boldsymbol{\varepsilon}_t$ are i.i.d.

(ii) $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \underset{r \times r}{\Sigma_{\varepsilon}}$

(iii) $\sum_{j=0}^{\infty} \|B_j\| < \infty$ where $\|B_j\|^2 = tr(B_j' B_j)$ and $B_0 = I_r$

(iv) $\det\{B(z)\} \neq 0$ for $|z| \leq 1$ where $B(z) = \sum_{j=0}^{\infty} B_j z^j$

then the process in (17) can also be written as:

$$\mathbf{y}_t = \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t \tag{18}$$

such that,

(v) $\sum_{j=1}^{\infty} \|A_j\| < \infty$

(vi) $A(z) = I_r - \sum_{j=1}^{\infty} A_j z^j = B(z)^{-1}$

(vii) $\det\{A(z)\} \neq 0$ for $|z| \leq 1$.

Jordà's (2005) local projection method of estimating the impulse response function is based instead on the expression that results from simple recursive substitution in the $VAR(\infty)$ representation, that is

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + \boldsymbol{\varepsilon}_{t+h} + B_1 \boldsymbol{\varepsilon}_{t+h-1} + ... + B_{h-1} \boldsymbol{\varepsilon}_{t+1} \tag{19}$$

where:

(i) $A_1^h = B_h$ for $h \geq 1$

(ii) $A_j^h = B_{h-1} A_j + A_{j+1}^{h-1}$ where $h \geq 1$; $A_{j+1}^0 = 0$; $B_0 = I_r$; and $j \geq 1$.

Now consider truncating the infinite lag expression (19) at lag k

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + A_k^h \mathbf{y}_{t-k+1} + \mathbf{v}_{k,t+h} \tag{20}$$

$$\mathbf{v}_{k,t+h} = \sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j}.$$

In what follows, we show that least squares estimates of (20) produce consistent estimates for A_j^h for $j = 1, ..., k$, in particular A_1^h, which is a direct estimate of the impulse response coefficient B_h. We obtain many of the derivations that follow by building on the results in Lewis and Reinsel (1985), who show that the coefficients of a truncated $VAR(\infty)$ are asymptotically normal as long as the truncation lag grows with the sample size at an appropriate rate.

Let $\Gamma(j) \equiv E(\mathbf{y}_t \mathbf{y}_{t+j}')$ with $\Gamma(-j) = \Gamma(j)'$. Further define:

(i) $X_{t,k} = \left(\mathbf{y}'_t, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-k+1}\right)'$ that is, the regressors in (20).

(ii) $\underset{kr\times r}{\widehat{\Gamma}_{1,k,h}} = (T-k-h)^{-1} \sum_{t=k}^{T-h} X_{t,k}\mathbf{y}'_{t+h}$

(iii) $\widehat{\Gamma}_k = (T-k-h)^{-1} \sum_{t=k}^{T-h} X_{t,k}X'_{t,k}$

Then, the mean-square error linear predictor of $\mathbf{y}_{t+h}$ based on $\mathbf{y}_t, ..., \mathbf{y}_{t-k+1}$ is given by the least-squares formula

$$\underset{r\times kr}{\widehat{A}}(k,h) = (\widehat{A}_1^h, ..., \widehat{A}_k^h) = \widehat{\Gamma}'_{1,k,h}\widehat{\Gamma}_k^{-1} \tag{21}$$

The following theorem establishes the consistency of these least-squares estimates for $A(k,h) = \left(A_1^h, ..., A_k^h\right)$.

Theorem 1 ***Consistency.*** *Let $\{y_t\}$ satisfy (17) and assume that:*

(i) $E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| < \infty$ for 1$\leq i,j,k,l \leq r$

(ii) k is chosen as a function of T such that

$$\frac{k^2}{T} \to 0 \text{ as } T, k \to \infty$$

(iii) k is chosen as a function of T such that

$$k^{1/2} \sum_{j=k+1}^{\infty} \|A_j\| \to 0 \text{ as } T, k \to \infty$$

Then:

$$\left\|\widehat{A}(k,h) - A(k,h)\right\| \xrightarrow{p} 0$$

The proof of this theorem is in the appendix. A natural consequence of the theorem provides the essential result, namely $\widehat{A}_1^h \xrightarrow{p} B_h$.

3.2 Asymptotic Normality

We now show that least-squares estimates from the truncated projections in (20) are asymptotically normal, although for the purposes of the PMD estimator, proving that $\widehat{A}_1^h$ is asymptotically normally distributed would suffice. Notice that we can write

$$\widehat{A}(k,h) - A(k,h) = \left\{ (T-k-h)^{-1} \sum_{t=k}^{T-h} \mathbf{v}_{k,t+h} X'_{t,k} \right\} \widehat{\Gamma}_k^{-1}$$

$$= (T-k-h)^{-1} \left[\sum_{t=k}^{T-h} \left\{ \left(\sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} \right) + \boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j} \right\} X'_{t,k} \right] \widehat{\Gamma}_k^{-1}$$

$$= (T-k-h)^{-1} \left\{ \sum_{t=k}^{T-h} \left(\sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} \right) X'_{t,k} \right\} \left\{ \Gamma_k^{-1} + \left(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1} \right) \right\} +$$

$$(T-k-h)^{-1} \left\{ \sum_{t=k}^{T-h} \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j} \right) X'_{t,k} \right\} \left\{ \Gamma_k^{-1} + \left(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1} \right) \right\}$$

Hence, the strategy of the proof will consist in showing that the first term in the sum above vanishes in probability and that the second term converges in probability as follows,

$$(T-k-h)^{1/2}\, vec \left[\widehat{A}(k,h) - A(k,h) \right] \xrightarrow{p}$$

$$(T-k-h)^{1/2}\, vec \left[(T-k-h)^{-1} \left\{ \sum_{t=k}^{T-h} \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j} \right) X'_{t,k} \right\} \Gamma_k^{-1} \right]$$

so that by showing that this last term is asymptotically normal, we complete the proof.

Define,

$$U_{1T} = \left\{ (T-k-h)^{-1} \sum_{t=k}^{T-h} \left(\sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} \right) X'_{t,k} \right\}$$

$$U_{2T}^* = \left\{ (T-k-h)^{-1} \sum_{t=k}^{T-h} \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j} \right) X'_{t,k} \right\}$$

then

$$(T-k-h)^{1/2}\, vec \left[\widehat{A}(k,h) - A(k,h) \right] =$$

$$(T-k-h)^{1/2} \left\{ \begin{array}{c} vec\left[U_{1T} \Gamma_k^{-1} \right] + vec \left[U_{1T} \left(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1} \right) \right] \\ + vec\left[U_{2T}^* \Gamma_k^{-1} \right] + vec \left[U_{2T}^* \left(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1} \right) \right] \end{array} \right\}$$

hence

$$(T-k-h)^{1/2}\, vec\left[\widehat{A}(k,h)-A(k,h)\right]-(T-k-h)^{1/2}\, vec\left[U_{2T}^{*}\Gamma_k^{-1}\right]=$$

$$(T-k-h)^{1/2}\left\{\begin{array}{c} vec\left[U_{1T}\Gamma_k^{-1}\right]+vec\left[U_{1T}\left(\widehat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\right] \\ +vec\left[U_{2T}^{*}\left(\widehat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\right]\end{array}\right\}=$$

$$\left(\Gamma_k^{-1}\otimes I_r\right) vec\left[(T-k-h)^{1/2}\, U_{1T}\right]+$$

$$\left\{\left(\widehat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\otimes I_r\right\} vec\left[(T-k-h)^{1/2}\, U_{1T}\right]+$$

$$\left\{\left(\widehat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\otimes I_r\right\} vec\left[(T-k-h)^{1/2}\, U_{2T}^{*}\right]$$

Define, with a slight change in the order of the summands,

$$\begin{aligned} W_{1T} &= \left\{\left(\widehat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\otimes I_r\right\} vec\left[(T-k-h)^{1/2}\, U_{1T}\right] \\ W_{2T} &= \left\{\left(\widehat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\otimes I_r\right\} vec\left[(T-k-h)^{1/2}\, U_{2T}^{*}\right] \\ W_{3T} &= \left(\Gamma_k^{-1}\otimes I_r\right) vec\left[(T-k-h)^{1/2}\, U_{1T}\right] \end{aligned}$$

then, in the next theorem we show that $W_{1T}\xrightarrow{p}0$, $W_{2T}\xrightarrow{p}0$, $W_{3T}\xrightarrow{p}0$.

Theorem 2 *Let $\{\mathbf{y}_t\}$ satisfy (17) and assume that*

(i) $E\left|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}\right|<\infty;\ 1\leq i,j,k,l\leq r$

(ii) k is chosen as a function of T such that $\frac{k^3}{T}\rightarrow 0,\ k,T\rightarrow\infty$

(iii) k is chosen as a function of T such that

$$(T-k-h)^{1/2}\sum_{j=k+1}^{\infty}\|A_j\|\rightarrow 0;\ \ k,T\rightarrow\infty$$

Then

$$(T-k-h)^{1/2}\, vec\left[\widehat{A}(k,h)-A(k,h)\right]\xrightarrow{p}$$

$$(T-k-h)^{1/2}\, vec\left[\left\{(T-k-h)^{-1}\sum_{t=k}^{T-h}\left(\boldsymbol{\varepsilon}_{t+h}+\sum_{j=1}^{h-1}B_j\boldsymbol{\varepsilon}_{t+h-j}\right)X_{t,k}'\right\}\Gamma_k^{-1}\right]$$

The proof is provided in the appendix. Now that we have shown that W_{1T}, W_{2T}, and W_{3T} vanish in probability, all that remains is to show that

$$A_T \equiv (T-k-h)^{1/2}\, vec\left[(T-k-h)^{-1}\left\{\sum_{t=k}^{T-h}\left(\boldsymbol{\varepsilon}_{t+h}+\sum_{j=1}^{h-1}B_j\boldsymbol{\varepsilon}_{t+h-j}\right)X_{t,k}^{'}\right\}\Gamma_k^{-1}\right] \xrightarrow{d}$$

$$N(0,\Omega_A) \text{ with } \Omega_A = \left(\Gamma_k^{-1}\otimes\Sigma_h\right);\ \Sigma_h = \left(\Sigma_\varepsilon + \sum_{j=1}^{h-1}B_j\Sigma_\varepsilon B_j'\right)$$

Since, $vec\left[\widehat{A}(k,h)-A(k,h)\right] \xrightarrow{p} A_T$, and $A_T \xrightarrow{d} N(0,\Omega_A)$, then we will have $vec\left[\widehat{A}(k,h)-A(k,h)\right] \xrightarrow{d} N(0,\Omega_A)$. We establish this result in the next theorem.

Theorem 3 *Let* $\{\mathbf{y}_t\}$ *satisfy (17) and assume*

(i) $E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| < \infty$*;* $1\leq i,j,k,l \leq r$

(ii) k is chosen as a function of T such that

$$\frac{k^3}{T} \to 0, k, T \to \infty$$

Then

$$A_T \xrightarrow{d} N(0,\Omega_A)$$

The proof is provided in the appendix.

In practice, we find it convenient to estimate responses for horizons 1, ..., h jointly as follows. Define,

(i) $\underset{r(k-1)\times 1}{X_{t-1,k}} \equiv \left(\mathbf{1}', \mathbf{y}_{t-1}', ..., \mathbf{y}_{t-k+1}'\right)'$ where $\mathbf{1}$ is a vector of ones for the constant term.

(ii) $\underset{rh\times 1}{Y_{t,h}} \equiv \left(\mathbf{y}_{t+1}', ..., \mathbf{y}_{t+h}'\right)'$

(iii) $\underset{1\times 1}{M_{t-1,k}} \equiv 1 - \sum_{t=k}^{T-h} X_{t-1,k}'\left(\sum_{t=k}^{T-h} X_{t-1,k}X_{t-1,k}'\right)^{-1} X_{t-1,k}$

(iv) $\underset{r\times r}{\widehat{\Gamma}_{1|k}} \equiv (T-k-h)^{-1}\sum_{t=k}^{T-h}\mathbf{y}_t M_{t-1,k}\mathbf{y}_t'$

(v) $\underset{r\times rh}{\widehat{\Gamma}_{1,h|k}} \equiv (T-k-h)^{-1}\sum_{t=k}^{T-h}\mathbf{y}_t M_{t-1,k} Y_{t,h}'$

Hence, the impulse response coefficient matrices for horizons 1 through h can be jointly estimated in a single step with

$$\widehat{\Gamma}'_{1,h|k}\widehat{\Gamma}^{-1}_{1|k} = \begin{bmatrix} \widehat{A}_1^1 \\ \widehat{A}_1^2 \\ \\ \widehat{A}_1^h \end{bmatrix} = \begin{bmatrix} \widehat{B}_1 \\ \widehat{B}_2 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = \widehat{B}(1,h) \tag{22}$$

Using the usual least-squares formulas, notice that

$$\widehat{B}(1,h) = B(1,h) + \left\{ (T-k-h)^{-1} \sum_{t=k}^{T-h} \mathbf{y}_t M_{t-1,k} V'_{t,h} \right\}' \widehat{\Gamma}^{-1}_{1|k} + o_p(1) \tag{23}$$

where $V_{t,h} \equiv \left(\mathbf{v}'_{t+1}, ..., \mathbf{v}'_{t+h}\right)'$; $\mathbf{v}_{t+j} = \boldsymbol{\varepsilon}_{t+j} + B_1\boldsymbol{\varepsilon}_{t+j-1} + ... + B_{j-1}\boldsymbol{\varepsilon}_{t+1}$ for $j = 1, ..., h$ and the terms vanishing in probability in (23) involve the terms U_{1T}, U_{2T}, and U_{3T} defined in the proof of theorem one, which makes use of the condition $k^{1/2}\sum_{j=k+1}^{\infty} ||A_j|| \to 0$ as $T, k \to \infty$. Under the conditions of theorem 2, we can write

$$\begin{aligned} &(T-k-h)^{1/2} vec\left(\widehat{B}(1,h) - B(1,h)\right) \xrightarrow{p} \\ &(T-k-h)^{1/2}\, vec\left[\left\{(T-k-h)^{-1}\sum_{t=k}^{T-h} V_{t,h} M_{t-1,k}\mathbf{y}'_t\right\}\widehat{\Gamma}^{-1}_{1|k}\right] \end{aligned} \tag{24}$$

from which we can derive the asymptotic distribution under theorems 2 and 3.

Next notice that

$$(T-k-h)^{-1}\sum_{t=k}^{T-h} V_{t,h}V'_{t,h} \xrightarrow{p} \underset{rh\times rh}{\Sigma_v} \tag{25}$$

The specific form of the variance-covariance matrix Σ_v can be derived as follows. Let $\mathbf{0}_j = \underset{j\times j}{\mathbf{0}}$; $\mathbf{0}_{m,n} = \underset{m\times n}{\mathbf{0}}$;and recall that $V_{t,h} \equiv \left(\mathbf{v}'_{t+1}, ..., \mathbf{v}'_{t+h}\right)'$, specifically,

$$V_{t,h} = \begin{bmatrix} \boldsymbol{\varepsilon}_{t+1} \\ \boldsymbol{\varepsilon}_{t+2} + B_1\boldsymbol{\varepsilon}_{t+1} \\ \vdots \\ \boldsymbol{\varepsilon}_{t+h} + B_1\boldsymbol{\varepsilon}_{t+h-1} + ... + B_{h-1}\boldsymbol{\varepsilon}_{t+1} \end{bmatrix} = \Psi_B \boldsymbol{\varepsilon}_{t,h},$$

where

$$\underset{rh\times rh}{\Psi_B} = \begin{bmatrix} I_r & \mathbf{0} & ... & \mathbf{0} \\ B_1 & I_r & ... & \mathbf{0} \\ \vdots & \vdots & ... & \vdots \\ B_{h-1} & B_{h-2} & ... & I_r \end{bmatrix}; \underset{rh\times 1}{\boldsymbol{\varepsilon}_{t,h}} = \begin{bmatrix} \boldsymbol{\varepsilon}_{t+1} \\ \vdots \\ \boldsymbol{\varepsilon}_{t+h} \end{bmatrix} \tag{26}$$

Then $E\left[V_{t,h}V'_{t,h}\right] = E\left[\Psi_B\boldsymbol{\varepsilon}_{t,h}\boldsymbol{\varepsilon}'_{t,h}\Psi'_B\right] = \Psi_B E\left[\boldsymbol{\varepsilon}_{t,h}\boldsymbol{\varepsilon}'_{t,h}\right]\Psi'_B$ with $E\left[\boldsymbol{\varepsilon}_{t,h}\boldsymbol{\varepsilon}'_{t,h}\right] = (I_h \otimes \Sigma_\varepsilon)$ and hence

$$E\left[V_{t,h}V'_{t,h}\right] = \Sigma_v = \Psi_B\left(I_h \otimes \Sigma_\varepsilon\right)\Psi'_B$$

and therefore

$$(T-k-h)^{1/2}\, vec\left(\widehat{B}(1,h) - B(1,h)\right) \xrightarrow{d} N\left(\mathbf{0},\Omega_B\right)$$

$$\underset{r^2h\times r^2h}{\Omega_B} = \left(\underset{r\times r}{\Gamma^{-1}_{1|k}} \otimes \underset{rh\times rh}{\Sigma_v}\right)$$

In practice, one requires sample estimates $\widehat{\Gamma}^{-1}_{1|k}$ and $\widehat{\Sigma}_v$. With respect to the latter, notice that the parametric form of expression (**??**) allows us to construct a sample estimate of Ω_B by plugging-in the estimates $\widehat{B}(1,h)$ and $\widehat{\Sigma}_\varepsilon$ into the expression (**??**).

3.3 Practical Summary of Results in Matrix Algebra

Define $\mathbf{y}_j$ for $j = h$, ..., 1, 0, –1, ..., –k as the $(T-k-h)\times r$ matrix of stacked observations of the $1\times r$ vector y'_{t+j}. Additionally, define the $(T-k-h)\times r(h+1)$ matrix $Y \equiv (\mathbf{y}_0, ..., \mathbf{y}_h)$; the $(T-k-h)\times r$ matrix $X \equiv \mathbf{y}_0$; the $(T-k-h)\times r(k-1)+1$ matrix $Z \equiv \left(\mathbf{1}_{(T-k-h)\times 1}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k+1}\right)$

and the $(T-k-h)\times(T-k-h)$ matrix $M_z = I_{T-k-h} - Z\left(Z'Z\right)^{-1}Z'$. Notice that the inclusion of $\mathbf{y}_0$ in Y is a computational trick that has no other effect but to ensure that the first block of coefficients is I_r, as is required for the minimum chi-square step. Using standard properties of least-squares

$$\widehat{B_T}(0,h) = \begin{bmatrix} I_r \\ \widehat{B}_1 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = \left[Y'M_zX\right]\left[X'M_zX\right]^{-1} \tag{27}$$

with an asymptotic variance-covariance matrix for $\widehat{b}_T = vec(\widehat{B_T}(0,h))$, that can be estimated with $\widehat{\Omega}_B = \left\{\left[X'M_zX\right]^{-1} \otimes \widehat{\Sigma}_v\right\}$. Properly speaking, the equations associated with $B_0 = I_r$ have zero variance, however, we find it notationally more compact and mathematically equivalent to calculate the residual variance-covariance matrix as $\widehat{\Sigma}_v = \widehat{\Psi}_B\left(I_{h+1} \otimes \widehat{\Sigma}_\epsilon\right)\widehat{\Psi}'_B$, and by extending $\widehat{\Psi}_B$ in (26) as

$$\underset{r(h+1)\times r(h+1)}{\widehat{\Psi}_B} = \begin{bmatrix} \mathbf{0}_r & \mathbf{0}_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ \mathbf{0}_r & I_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ \mathbf{0}_r & \widehat{B}_1 & I_r & ... & \mathbf{0}_r \\ \vdots & \vdots & \vdots & ... & \vdots \\ \mathbf{0}_r & \widehat{B}_{h-1} & \widehat{B}_{h-2} & ... & I_r \end{bmatrix} \tag{28}$$

with $\widehat{B}_j$ replacing B_j, $\widehat{\Sigma}_\epsilon = \frac{\widehat{\mathbf{v}_1}'\widehat{\mathbf{v}_1}}{T-k-h}$; $\widehat{\mathbf{v}_1} = M_z y_1 - M_z\mathbf{y}_0\widehat{B}_1$.

4 The Second Step: Minimum Chi-Square

We return now to deriving the statistical properties of the estimator resulting from minimization of expression (16). The section begins by deriving consistency and asymptotic normality, it shows how contemporaneous restrictions expand these basic results and then derives an overall test of model misspecification based on overidentifying restrictions. The section concludes with a summary of the main results for practitioners.

4.1 Consistency

Given an estimate of $\mathbf{B}$ (and hence b) from the first-stage described in the previous section, our objective here is to estimate $\phi = \{vec(\Phi_1)\ vec(\Phi_2)\}'$ by minimizing

$$\min_{\phi} \widehat{Q}_T(\phi) = \left[S\widehat{b}_T - g\left(\widehat{b}_T;\phi\right)\right]' \widehat{W}\left[S\widehat{b}_T - g\left(\widehat{b}_T;\phi\right)\right]$$

where the reader is reminded that

$$\begin{aligned} S &\equiv (R' \otimes S_0) \\ g\left(\widehat{b}_T;\phi\right) &\equiv \left\{\left(I_{r_1} \otimes S_1\widehat{\mathbf{B}}_T R\right) \quad \left(I_{r_1} \otimes S_2\widehat{\mathbf{B}}_T R\right)\right\}\phi \end{aligned}$$

Let $Q_0(\phi)$ denote the objective function at b_0. The following theorem establishes the conditions under which $\widehat{\phi}_T$, the solution of the minimization problem, is consistent for ϕ_0.

Theorem 4 *Given that $\widehat{b}_T \xrightarrow{p} b_0$, assume that*

(i) $\widehat{W} \xrightarrow{p} W$*, a positive semidefinite matrix*

(ii) $Q_0(\phi)$ *is uniquely maximized at* ϕ_0

(iii) The parameter space Θ *is compact*

(iv) $Q_0(\phi)$ *is continuous*

(v) $(h-1)\left(r_1^2 + r_1 r_2\right) \geq \dim(\phi)$

Then

$$\widehat{\phi}_T \xrightarrow{p} \phi_0$$

The proof is provided in the appendix and consists of showing that $\widehat{Q}_T(\phi) \xrightarrow{p} Q_0(\phi)$ uniformly and that $\widehat{Q}_T(\phi)$ is stochastically equicontinuous. Next, we show that the minimum chi-square estimator is asymptotically normal.

4.2 Asymptotic Normality

The proof of asymptotic normality in classical minimum-distance estimation – where, as an example, $\widehat{b}_T = g(\phi)$, is a known, continuously differentiable function, and $\sqrt{T}\left(\widehat{b}_T - b_0\right) \xrightarrow{d} N\left(0, \Omega_B\right)$ – is rather straightforward. All that is required to obtain the distribution of $\widehat{\phi}_T$ is $G_\phi = \nabla_\phi g\left(\phi\right)$ to conclude that $\sqrt{T}\left(\widehat{\phi}_T - \phi\right) \xrightarrow{d} N\left(0, \left(G'_\phi \Omega_B^{-1} G_\phi\right)^{-1}\right)$, under mild regularity conditions.

Derivation of the distribution of PMD would be equivalent to the classical minimum-distance proof if we had $S\widehat{b}_T = g\left(b_0; \widehat{\phi}_T\right)$ instead of $S\widehat{b}_T = g\left(\widehat{b}_T; \widehat{\phi}_T\right)$. Although we know $\widehat{b}_T \xrightarrow{p} b_0$, $\widehat{b}_T$ is stochastic in finite samples and hence $g\left(\widehat{b}_T; \widehat{\phi}_T\right)$ is not, strictly speaking, a known function. Thus, derivation of the asymptotic distribution of $\widehat{\phi}_T$ requires that its asymptotic covariance matrix appropriately reflect the additional uncertainty in $g\left(\widehat{b}_T; \widehat{\phi}_T\right)$ and that conditions exist so that $g\left(\widehat{b}_T; \widehat{\phi}_T\right) \xrightarrow{p} g\left(b_0; \widehat{\phi}_T\right)$. These are the essential elements of theorem 5 below.

Theorem 5 *Given the following conditions:*

(i) $\widehat{W} \xrightarrow{p} W$*, a positive semidefinite matrix, where we choose* $W = \left(S\Omega_B S'\right)^{-1}$

(ii) $\widehat{b}_T \xrightarrow{p} b_0$ *and* $\widehat{\phi}_T \xrightarrow{p} \phi_0$ *from theorems 1 and 4.*

(iii) b_0 *and* ϕ_0 *are in the interior of their parameter spaces*

(iv) $g(\widehat{b}_T; \phi)$ *is continuously differentiable in a neighborhood* $\mathfrak{N}$ *of* θ_0, $\theta = \begin{pmatrix} b' & \phi' \end{pmatrix}'$

(v) $\sqrt{T}\left[S\widehat{b}_T - g(b_0; \phi_0)\right] \xrightarrow{d} N(0, S\Omega_B S')$.

(vi) There is a G_b *and* G_ϕ *that are continuous at* b_0 *and* ϕ_0 *respectively and*

$$\sup_{b,\phi \in \mathfrak{N}} \left\| \nabla_b g(b;\phi) - G_b \right\| \xrightarrow{p} \mathbf{0}$$

$$\sup_{b,\phi \in \mathfrak{N}} \left\| \nabla_\phi g(b;\phi) - G_\phi \right\| \xrightarrow{p} \mathbf{0}$$

(vii) For $G_\phi = G_\phi(\phi_0)$*, then* $G'_\phi W G_\phi$ *is invertible.*

(viii) $(h-1)\left(r_1^2 + r_1 r_2\right) \geq \dim(\phi)$

Then:

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} N\left(0, \Omega_\phi\right)$$

where

$$
\begin{aligned}
\Omega_\phi &= \left(G_\phi' W G_\phi\right)^{-1} + \\
&\left(G_\phi' W G_\phi\right)^{-1} G_\phi' W G_b \Omega_B G_b' W G_\phi \left(G_\phi' W G_\phi\right)^{-1} - \\
&\left(G_\phi' W G_\phi\right)^{-1} \left(G_\phi' W S \Omega_B G_b' W G_\phi\right) \left(G_\phi' W G_\phi\right)^{-1} - \\
&\left(G_\phi' W G_\phi\right)^{-1} \left(G_\phi' W G_b \Omega_B S' W G_\phi\right) \left(G_\phi' W G_\phi\right)
\end{aligned} \tag{29}
$$

The proof is provided in the appendix and essentially consists of applying the mean value theorem to the first order conditions of the minimization problem. Several results deserve comment. First, we derive the asymptotic covariance of $\widehat{\phi}_T$ by using the optimal weighting matrix, which in this case is $W = (S\Omega_B S')^{-1}$. Alternative weighting matrices are permissible and the appendix provides the general formula to calculate the appropriate asymptotic covariance matrix. Second, the expression for Ω_ϕ in (29) is the sum of four terms. The first is the expression of the asymptotic covariance matrix in classical minimum-distance. The remaining terms reflect the contribution to the variance of ϕ coming from the uncertainty of $\widehat{b}_T$ in $g(\widehat{b}_T; \phi)$.

The next section extends the weighted, quadratic, minimum-distance function to include the contemporaneous parameter conditions (5) we have so far set aside to provide the general result.

4.3 Incorporating Contemporaneous Parameter Restrictions

Up to this point, the derivation of the two-step PMD estimator has set aside the set of conditions

$$
P_1^{-1} = \Phi_0' I_{r_1} + \Phi_1' \mathbf{0} - \Phi_2' B_{11}^{1'}
$$

where $P_1 P_1' = \Sigma_{11} = E\left(\boldsymbol{\varepsilon}_{1t}\boldsymbol{\varepsilon}_{1t}'\right)$. Often, macroeconomic models specify $\Phi_0 = I_{r_1}$ and we will maintain this assumption in the discussion that follows to simplify our derivations, although we expect the reader will have no problem in extending our results otherwise. These contemporaneous conditions may be important in achieving identification in some models.

We can recast the previous expression to better match the stacked conditions in (12) by noticing that $P_1 = R'PR$ where $(\Sigma_\varepsilon = PP')$ as follows

$$\left[(R'PR)^{-1} - I\right] = (\mathbf{0B}R \quad S_{02}\mathbf{B}R)\begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} \tag{30}$$

where S_{02} is the selector matrix

$$S_{02} = \left[\mathbf{0}_r \quad I_r \quad \mathbf{0}_{r,r(h-1)}\right]$$

Applying the vec operator to both sides of expression (30) and letting

$$\begin{aligned} q &= vec(Q) \equiv vec\left((R'PR)^{-1} - I\right) \\ f\left(\widehat{b}_T;\phi\right) &\equiv \left(I_{r_1} \otimes (\mathbf{0B}R \quad S_{02}\mathbf{B}R)\right)\phi \end{aligned} \tag{31}$$

then, the sample vector expression of (30) is

$$\widehat{q}_T = f\left(\widehat{b}_T;\phi\right)$$

where P can be estimated from the Cholesky decomposition of the residual covariance matrix $\widehat{\Sigma}_\varepsilon$ to obtain $\widehat{q}_T = vec\left(\left(R'\widehat{P}_T R\right)^{-1} - I\right)$. The expressions (31) allow us to recast the minimum-distance of the second step in PMD as

$$\begin{bmatrix} \widehat{q}_T \\ S\widehat{b}_T \end{bmatrix} = \begin{bmatrix} f\left(\widehat{b}_T;\phi\right) \\ g\left(\widehat{b}_T;\phi\right) \end{bmatrix}$$

with the minimum-distance objective function

$$\min_{\phi} \widehat{Q}_T(\phi) = \begin{bmatrix} \widehat{q}_T - f\left(\widehat{b}_T;\phi\right) \\ S\widehat{b}_T - g\left(\widehat{b}_T;\phi\right) \end{bmatrix}' \widehat{W}_q \begin{bmatrix} \widehat{q}_T - f\left(\widehat{b}_T;\phi\right) \\ S\widehat{b}_T - g\left(\widehat{b}_T;\phi\right) \end{bmatrix}$$

and where

$$\widehat{W}_q = \begin{bmatrix} \widehat{\Omega}_q^{-1} & 0 \\ 0 & \left(S\widehat{\Omega}_B S'\right)^{-1} \end{bmatrix}$$

since we note that the covariance between the Cholesky decomposition of $\widehat{\Sigma}_\varepsilon$ and the impulse response function $\widehat{\mathbf{B}}_T$ is zero (see Lütkepohl, 1993). The following theorem establishes the asymptotic distribution of $\widehat{\phi}_T$ for this extended estimator.

Theorem 6 *Given the following conditions:*

(i) $\widehat{W}_q \xrightarrow{p} W_q$, *a positive semidefinite matrix,* $W_q = \begin{bmatrix} \Omega_q^{-1} & 0 \\ 0 & (S\Omega_B S')^{-1} \end{bmatrix}$, *where* Ω_q *is the variance of* $\widehat{q}_T$

(ii) $\widehat{q}_T \xrightarrow{p} q_0, \widehat{b}_T \xrightarrow{p} b_0$ *and* $\widehat{\phi}_T \xrightarrow{p} \phi_0$ *where the first two conditions follow from Theorem 1 and the second from Theorem 4.*

(iii) q_0, b_0 *and* ϕ_0 *are in the interior of their parameter spaces*

(iv) $f(\widehat{b}_T;\phi)$ *and* $g(\widehat{b}_T;\phi)$ *are continuously differentiable in a neighborhood* $\mathfrak{N}$ *of* θ_0, $\theta = \left(b' \quad \phi'\right)'$

(v) $\sqrt{T}\left(\widehat{q}_T - q_0\right) \xrightarrow{d} N\left(0, \Omega_q\right); \sqrt{T}\left[S\widehat{b}_T - g(b_0;\phi_0)\right] \xrightarrow{d} N(0, S\Omega_B S')$ *and* $E\left(\left(\widehat{q}_T - q_0\right), \left(S\widehat{b}_T - g(b_0;\phi_0)\right)\right) = 0$, *which are a consequence of Theorem 3.*

(vi) There is a F_b, F_ϕ, G_b *and* G_ϕ *that are continuous at* b_0 *and* ϕ_0 *respectively and*

$$\sup_{b,\phi\in\mathfrak{N}} \|\nabla_b f(b;\phi) - F_b\| \xrightarrow{p} \mathbf{0}$$

$$\sup_{b,\phi\in\mathfrak{N}} \|\nabla_\phi f(b;\phi) - F_\phi\| \xrightarrow{p} \mathbf{0}$$

$$\sup_{b,\phi\in\mathfrak{N}} \|\nabla_b g(b;\phi) - G_b\| \xrightarrow{p} \mathbf{0}$$

$$\sup_{b,\phi\in\mathfrak{N}} \|\nabla_\phi g(b;\phi) - G_\phi\| \xrightarrow{p} \mathbf{0}$$

(vii) For $H_\phi = \left(F_\phi\left(\phi_0\right) \quad G_\phi(\phi_0)\right)'$, *then* $H_\phi' W_q H_\phi$ *is invertible.*

(viii) $h\left(r_1^2 + r_1 r_2\right) \geq \dim(\phi)$

Then

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} N\left(0, V_\phi\right)$$

where

$$
\begin{aligned}
V_\phi &= \left(H_\phi' W_q H_\phi\right)^{-1} + \\
& \left(H_\phi' W_q H_\phi\right)^{-1} H_\phi' W_q H_b \Omega_B H_b' W_q H_\phi \left(H_\phi' W_q H_\phi\right)^{-1} - \\
& \left(H_\phi' W_q H_\phi\right)^{-1} H_\phi' W_q S \Omega_B H_b' W_q H_\phi \left(H_\phi' W_q H_\phi\right)^{-1} - \\
& \left(H_\phi' W_q H_\phi\right)^{-1} H_\phi' W_q H_b \Omega_B S' W_q H_\phi \left(H_\phi' W_q H_\phi\right)^{-1}
\end{aligned}
$$

with $H_\phi = (F_\phi(\phi_0) \quad G_\phi(\phi_0))'$, $H_b = (F_b(\phi_0) \quad G_b(\phi_0))'$. Specifically,

$$
\begin{aligned}
H_\phi &= \begin{pmatrix} I_{r_1} \otimes (\mathbf{0}\mathbf{B}R \quad S_{02}\mathbf{B}R) \\ I_{r_1} \otimes (S_1\mathbf{B}R \quad S_2\mathbf{B}R) \end{pmatrix}, \\
H_b &= [(\Phi_1' \quad \Phi_2') \otimes I_h] \begin{pmatrix} R' \otimes \mathbf{0} \\ R' \otimes S_{02} \\ R' \otimes S_1 \\ R' \otimes S_2 \end{pmatrix}
\end{aligned}
$$

The proof is provided in the appendix and parallels the proof in theorem 5. For completeness, we also report here the formula for Ω_q which is shown in the appendix to be

$$
\begin{aligned}
\Omega_q &= 2\Gamma D_r^+ (\Sigma_\varepsilon \otimes \Sigma_\varepsilon) D_r^{+'} \Gamma' \\
\Gamma &= \left[(RP'R')^{-1} \otimes (R'PR)^{-1}\right] [R \otimes R'] L_r' \{L_r (I_{r^2} + K_{rr}) (P \otimes I_r) L_r'\}^{-1}
\end{aligned}
$$

were L_r is the elimination matrix such that, for any square, $r \times r$ matrix Σ then $vech(\Sigma) = L_r vec(\Sigma)$; K_{rr} is the commutation matrix such that $vec(\Sigma) = K_{rr} vec(\Sigma')$ and $D_r^+ = (D_r' D_r)^{-1} D_r$ where D_r is the duplication matrix such that $vec(\Sigma) = D_r vech(\Sigma)$.

4.4 Test of Overidentifying Restrictions

The second stage in PMD consists of minimizing a weighted quadratic distance to obtain estimates of the parameter vector ϕ, which contains $2r_1^2$ elements. The identification conditions require that the impulse response horizon h be chosen to guarantee that there are at least as many moment conditions as elements in ϕ. When the number of moment conditions coincides with the dimension of ϕ, the quadratic function $\widehat{Q}_T(\phi)$ obtains its lower bound of 0. However, when the number of conditions is larger than the dimension of ϕ, the lower bound 0 is only achieved if the model is correctly specified, as the sample size grows to infinity. This observation forms the basis of the test for overidentifying restrictions (or J-test) in GMM and is a feature that can be exploited to construct a similar test for PMD.

We begin by noting that under the conditions of theorems 5 and 6,

$$\begin{bmatrix} \widehat{q}_T - f(b_0) \\ S\widehat{b}_T - g(b_0; \phi_0) \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} \Omega_q & \mathbf{0} \\ \mathbf{0} & S\Omega_B S' \end{bmatrix}\right)$$

so that the minimum-distance function $\widehat{Q}_T\left(\widehat{\phi}_T\right)$ evaluated at the optimum is a quadratic form of standardized normally distributed random variables (since the optimal $\widehat{W}_q$ is the inverse of the variance in the previous expression) and therefore, distributed χ^2 with degrees of freedom $h\left(r_1^2 + r_1 r_2\right) - 2r_1^2$, or simply $\dim\left[\begin{pmatrix} \widehat{q}_T & S\widehat{b}_T \end{pmatrix}'\right] - \dim(\widehat{\phi}_T)$.

4.5 PMD: A Summary for Practitioners

Consider an economy characterized by an $r \times 1$ vector of variables $\mathbf{y}_t = (\mathbf{y}_{1t}' \quad \mathbf{y}_{2t}')'$ where $\mathbf{y}_{1t}$ and $\mathbf{y}_{2t}$ are sub-vectors of dimensions r_1 and r_2 respectively, with $r = r_1 + r_2$. A researcher specifies a macroeconomic model for the variables in $\mathbf{y}_{1t}$ whose Euler equations can be summarized as

$$\mathbf{y}_{1t} = \Phi_1' \mathbf{y}_{1t-1} + \Phi_2' E_t \mathbf{y}_{1t+1} + \mathbf{u}_{1t}, \quad E\left(\mathbf{u}_{1t}' \mathbf{u}_{1t}\right) = I$$

The following steps summarize how PMD can be used to estimate the parameters in Φ_1 and Φ_2 :

First Stage: Local Projections

1. Construct $Y = (\mathbf{y}_0, ..., \mathbf{y}_h)'$; $X = \mathbf{y}_0$; $Z = \left(\mathbf{1}_{(T-k-h)\times r}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k+1}\right)$; $M_z = I_{(T-k-h)} - Z\left(Z'Z\right)^{-1} Z$, where $\mathbf{y}_j$ is the $(T-k-h) \times r$ matrix of observations for the vector $\mathbf{y}_{t+j}$.

2. Compute by least squares $\widehat{b}_T = vec(\widehat{B}(0,h))$, where

$$\widehat{B}(0,h) = [Y'M_zX]\,[X'M_zX]^{-1}$$

3. Calculate the covariance matrix of b as $\widehat{\Omega}_B = \left\{(X'M_zX)^{-1} \otimes \widehat{\Sigma}_v\right\}$, where $\widehat{\Sigma}_v = \widehat{\Psi}_B \left(I_h \otimes \widehat{\Sigma}_\varepsilon\right) \widehat{\Psi}'_B$, $\widehat{\Psi}_B$ is given by expression (28), and $\widehat{\Sigma}_\varepsilon = (\widehat{\mathbf{v}}_1'\widehat{\mathbf{v}}_1) \,/\, (T-k-h)$; with $\widehat{\mathbf{v}}_1 = M_z\mathbf{y}_1 - M_z\mathbf{y}_0\widehat{B}_1$.

   Second Stage: Minimum Chi-Square

4. Recall the definitions of $S\widehat{b}_T$; $\widehat{H}_\phi$; and $\widehat{W}_q$. An estimate of ϕ can be obtained with weighted least-squares as

$$\widehat{\phi}_T = \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} \left(\widehat{H}'_\phi \widehat{W}_q S\widehat{b_T}\right)$$

5. The covariance matrix of $\widehat{\phi}_T$ can be estimated as

$$\begin{aligned} \widehat{V}_\phi = {} & \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} + \\ & \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} \widehat{H}'_\phi \widehat{W}_q \widehat{H}_b \widehat{\Omega}_B \widehat{H}'_b \widehat{W}_q \widehat{H}_\phi \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} - \\ & \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} \widehat{H}'_\phi \widehat{W}_q S \widehat{\Omega}_B \widehat{H}'_b \widehat{W}_q \widehat{H}_\phi \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} - \\ & \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} \widehat{H}'_\phi \widehat{W}_q \widehat{H}_b \widehat{\Omega}_B S' \widehat{W}_q \widehat{H}_\phi \left(\widehat{H}'_\phi \widehat{W}_q \widehat{H}_\phi\right)^{-1} \end{aligned}$$

6. And a test of model misspecification can be constructed as

$$\widehat{Q}_T\left(\widehat{\phi}_T\right) \xrightarrow{d} \chi^2_{h\left(r_1^2 + r_1 r_2\right) - 2r_1^2}$$

5 The Relation between GMM and PMD: An Example

PMD and GMM are both minimum distance methods. In this section we use a simple motivating example to compare the advantages of PMD over GMM and will show that GMM can be thought of as a special case of PMD. To keep things simple, suppose the DGP is characterized by the univariate backward/forward model:

$$y_t = \phi_1 y_{t-1} + \phi_2 E_t y_{t+1} + \varepsilon_t. \tag{32}$$

Instead, suppose the Euler condition from a proposed rational expectations model can be expressed as

$$y_t = \rho E_t y_{t+1} + u_t, \tag{33}$$

which is misspecified with respect to the DGP. Based on the economic model in (33), any y_{t-j}; $j > 1$ would be considered a valid instrument for GMM estimation and hence, an estimate of ρ would be found with the set of conditions

$$\widehat{\rho}_{GMM} = \left(\frac{1}{T} \sum^{T} y_{t-j} y_{t+1} \right)^{-1} \left(\frac{1}{T} \sum^{T} y_{t-j} y_t \right). \tag{34}$$

It is easy to see that the probability limit of these conditions is

$$\widehat{\rho}_{GMM} \xrightarrow{p} \phi_2 + \phi_1 \frac{\gamma_{j-1}}{\gamma_{j+1}}; j \geq 1$$

where $\gamma_j = COV(y_t y_{t-j})$. Notice that the bias, $\phi_1 \frac{\gamma_{j-1}}{\gamma_{j+1}}$, does not disappear by selecting longer lags of y_{t-j} as instruments, since although $\gamma_j \to 0$ as $j \to \infty$, $\frac{\gamma_{j-1}}{\gamma_{j+1}}$ is indeterminate as both the numerator and the denominator are simultaneously going to zero. Meanwhile, as $j \to \infty$ the correlation of the instrument with the regressor is exponentially decaying to zero – not only are these instruments invalid, they are increasingly weak. The validity of the instruments obviously depends on the dynamics of the DGP, not on the dynamics of the proposed economic model.

PMD takes on a more agnostic view on the dynamics of the DGP. The MA(∞) representation of (32) is

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$$

and hence, under the proposed model in (33), PMD would first estimate the b_j by local projections and then use the mapping between the b_j and the ρ implied by the proposed model, which in this simple case is:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_h \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ b_{h-1} \end{bmatrix} \rho \tag{35}$$

Local projections for the b_j are

$$\widehat{b}_j = \left(\frac{1}{T} \sum^{T} y'_{t-j} M_{t-j} y_{t-j} \right)^{-1} \left(\frac{1}{T} \sum^{T} y'_{t-j} M_{t-j} y_t \right) \tag{36}$$

where $M_t = 1 - z_t \left(z'_t z_t\right)^{-1} z_t$ and $z_t = (1 \; y_{t-2} \; ...y_{t-k+1})$. Notice that

$$b_j = \rho b_{j-1}$$

for $j \geq 1$ so that an estimate of ρ can be obtained directly from the local projections by noticing that the common term $\left(\frac{1}{T} \sum^{T} y'_{t-j} M_{t-j} y_{t-j} \right)^{-1}$ cancels out on both sides of expression (36) to obtain

$$\widehat{\rho}_{PMD} = \left(\frac{1}{T} \sum^{T} y'_{t-j} M_{t-j} y_{t+1} \right)^{-1} \left(\frac{1}{T} \sum^{T} y'_{t-j} M_{t-j} y_t \right) \tag{37}$$

which is the PMD counterpart to expression (34). However, notice that although the proposed model is misspecified with respect to the DGP, PMD delivers an unbiased estimate of the structural parameter of interest, that is

$$\widehat{\rho}_{PMD} \xrightarrow{p} \phi_2$$

In other words, PMD succeeds in consistently estimating the parameter ρ from the misspecified proposed economic model (33). What explains this surprising result? In practical terms and for this simple example only, PMD turns out to be equivalent to pre-treating the candidate instruments by conditioning either on past values of the variables and/or omitted variables (in a more general case), so that only the marginal information left after conditioning is used to instrument.

The first-stage local projections therefore serve to eliminate the sources of inconsistency in the instruments, which then enter the second stage estimation weighted by the relative strength of the conditional correlation with the instrumented variable. PMD resolves the appropriate asymptotic theory associated with this pre-treatment in an indirect way. On the other hand, GMM relies on finding valid instruments (in both the dynamic and the traditional sense) unconditionally in their raw form. Unfortunately, the proposed economic model usually offers insufficient guidance as to what these instruments might be.

6 Monte Carlo Experiments: Estimating ARMA(p,q) models with PMD

This section investigates the small sample properties of PMD. We take this opportunity to further demonstrate the flexibility of our method by experimenting with univariate $ARMA(1,1)$ specifications, which would typically require numerical optimization routines. However, we find there is pedagogical value in discussing the more general $VARMA(1,1)$ model so that the reader can readily generalize the method to $VARMA(p,q)$ specifications. Accordingly, let $\mathbf{y}_t$ be an $r \times 1$ vector that follows the following covariance-stationary process

$$\mathbf{y}_t = \underset{r\times r}{\Pi_1'}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t + \underset{r\times r}{\Theta_1'}\boldsymbol{\varepsilon}_{t-1} \tag{38}$$

with Wold decomposition,

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j' \boldsymbol{\varepsilon}_{t-j} \tag{39}$$

with $B_0 = I_r$. Substituting (39) into (38) and equating terms in $\boldsymbol{\varepsilon}_{t-j}$ the same way we did in section 2, we obtain the following conditions:

$$\begin{aligned} B_1' &= I_r\Pi_1' + \Theta_1' \\ B_j' &= B_{j-1}'\Pi_1' \quad \text{for } j > 1 \end{aligned} \tag{40}$$

Consider now stacking the first h of these conditions. To that end, modify the definition of the selector matrices introduced in section 2 as follows (the star serves to distinguish the definitions from those in previous sections):

$$\begin{aligned} S_0^* &= \left[\mathbf{0}_{rh,r} \quad (I_h \otimes I_r) \right]; \\ S_1^* &= \left[(I_h \otimes I_r) \quad \mathbf{0}_{rh,r} \right]; \\ S_2^* &= \begin{bmatrix} I_r & \mathbf{0}_{r,rh} \\ \multicolumn{2}{c}{\mathbf{0}_{r(h-1),r(h+1)}} \end{bmatrix}. \end{aligned} \tag{41}$$

Defining $\mathbf{B} = B(0,h)$ as in expression (7), then it should be clear that the conditions in (40) can be expressed as

$$S_0^* \mathbf{B} = S_1^* \mathbf{B} \Pi_1 + S_2^* \mathbf{B} \Theta_1$$

so that the associated minimum-distance function is totally analogous to expression (14), specifically,

$$\begin{gathered} (I_r \otimes S_0^*) \widehat{b}_T - \left(I_r \otimes \left(S_1^* \widehat{\mathbf{B}}_T \quad S_2^* \widehat{\mathbf{B}}_T \right) \right) \phi = \\ S^* \widehat{b}_T - g^* (\widehat{b}_T; \phi) \end{gathered} \tag{42}$$

where $\phi = vec\,(\Pi_1 \quad \Theta_1)$ and estimation consists in finding the solution to the problem

$$\min_{\lambda} \widehat{Q^*}_T (\phi) = \left[S^* \widehat{b}_T - g^* \left(\widehat{b}_T; \phi \right) \right]' \widehat{W^*} \left[S^* \widehat{b}_T - g^* \left(\widehat{b}_T; \phi \right) \right]$$

with $\widehat{W^*} = \left(S^* \widehat{\Omega}_B S^{*\prime} \right)^{-1}$. It should be immediately obvious that once one defines the new selector matrices (41), estimation of the parameters of the model and calculation of the standard errors can be done exactly as described in section 4.5.

The set-up of the Monte Carlo experiments is as follows. We investigate four different parameter pairs (π_1, θ_1) for the univariate ARMA(1,1) specification

$$y_t = \pi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

Specifically: cases (i) and (ii) are two ARMA(1,1) models with parameters (0.25, 0.5) and (0.5, 0.25) respectively, and cases (iii) and (iv) are a pure MA(1) model with parameters (0, 0.5) and a pure AR(1) model with parameters (0.5, 0), both estimated as general ARMA(1,1) models. In addition, we generated data from two AR(2) models

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \varepsilon_t$$

with parameter pairs (π_1, π_2) given by (0.5, 0.25) and (0.25, 0.5). We use the AR(2) models as a way to check the misspecification test based on the test of overidentifying restrictions. Thus, for the models with $\theta_1 = 0.5$ (cases (i) and (iii)), we use the alternative ($\pi_1 = 0.25$, $\pi_2 = 0.5$) and for the models with $\pi_1 = 0.5$ (cases (ii) and (iv)), we use the alternative ($\pi_1 = 0.5, \pi_2 = 0.25$). Clearly, the alternative model for cases (i) and (iii) has rather different dynamics than the original model whereas the alternative model in cases (ii) and (iv) is very similar to the original model. This design is meant to illustrate the relative power of the misspecification test.

Each simulation run has the following features. We use a burn-in of 500 observations that we then disregard to avoid initialization problems. We experiment with practical sample sizes $T = 50$, 100, and 400 observations. The lag-length of the first-stage PMD estimator is determined automatically by AIC_c.[2] For the second stage, we experimented with impulse response horizons $h = 2, 5,$ and 10. When $h = 2$, we have exact identification, otherwise, the model is overidentified. Although the impulse responses for the models we simulate decay within two to three periods, we experimented with $h = 10$ to examine the effects of including many additional conditions, that would seem not to include any useful information for parameter estimation.

The models in each of cases (i)-(iv) is estimated by both maximum likelihood (MLE) and

[2] AIC_c refers to the correction to AIC introduced in Hurvich and Tsai (1989), which is specifically designed for autoregressive models. There were no significant differences when using SIC or the traditional AIC.

PMD and we report Monte Carlo averages and standard errors of the parameter estimates, as well as Monte Carlo averages of standard error estimates based on the MLE and PMD formulas. The objective is to ensure that the coverage implied by the analytical formulas corresponds to the Monte Carlo coverage. Finally, we report two chi-square tests. The first test, labeled $\chi^2 - corr$ is a test of overidentifying restrictions when the model is correctly specified as an ARMA(1,1) under any of cases (i)-(iv). The second test is labeled $\chi^2 - incorr$ and is a test of overidentifying restrictions when the true model simulated is the AR(2) model described above but an ARMA(1,1) is specified instead. The Monte Carlo average of the p-value of the first test offers some guidance as to the size of the test whereas the Monte Carlo average of the p-value of the second test speaks to the power of the test depending on which of the two AR(2) models is used to simulate the alternative. Although a more comprehensive Monte Carlo on the properties of the test for overidentification is desirable, we felt this test is subsidiary to the estimation strategy that is the main thrust of the paper and leave for future research a more exhaustive exploration of its properties. Finally, we used 500 replications for each experiment.

Tables 1-4 contain the results for each of cases (i)-(iv). Several results deserve comment. First, PMD estimates converge to the true parameter values at roughly the same speed (sometimes faster) as MLE estimates, with estimates being close to the true values even in samples of 50 observations. However, with 50 observations, we remark some deterioration of PMD parameter estimates when $h = 10$, as would be expected by the loss of degrees of freedom. Second, PMD has slightly wider analytic standard errors than MLE. Notice that PMD achieves the MLE lower bound only asymptotically when $h \to \infty$ as $T \to \infty$. Hence when $T = 400$ and $h = 10$, examples of PMD/MLE standard errors are: 0.075/0.072, 0.066/0.064, 0.075/0.066, 0.077/0.072, 0.090/0.088, 0.099/0.100 (we omit case (iii) since MLE estimates are for a pure MA(1) specification instead). Third, we find that the analytical formula for the PMD standard errors provides similar and correct coverage to the analytical formula for MLE, both relative to their Monte Carlo standard errors. Fourth, the average p-value of the test of overidentifying restrictions when the model is correctly specified is approximately 0.50. In those cases where the true model is an AR(2) with parameter pair (0.25, 0.5), we found the test to correctly detect the misspecification with samples of 100 observations or more (with average p-values of 0.025 and below). The test had more difficulty in distinguishing the AR(2) model with parameter pair (0.50, 0.25), which was to be expected since this AR(2) model differs little from the ARMA(1,1) or case (ii) of the pure AR(1) with coefficient 0.5 of case (iv). Even with a sample size of 400 observations, the average p-value was still about 0.14.

Finally, we also remark that MLE estimates of the ARMA(1,1) specification for case (iii) in table 3 failed to converge due to numerical instability – the likelihood is nonlinear in the parameters and has to be optimized numerically. Hence, we report MLE results for a pure MA(1) specification. We faced a similar problem for case (iv) in table 4 and with a sample size $T = 50$ where we had to estimate pure AR(1) specifications. However, we had no problems for $T = 100$, and $T = 400$. Naturally, PMD does not suffer from these numerical approximation issues and hence we reported ARMA(1,1) specifications in all cases.

Summarizing, PMD performs very well in this set of experiments. We found that the optimal weighting matrix does a good job at appropriately bringing in information from impulse responses at long horizons that may be contaminated with significant sample variation. In our experiments, parameter estimates are very stable to the choice of horizon h, the only consequence being an expected reduction in standard errors. Naturally, this statement depends on the sample size and hence the degrees of freedom available for the first-stage estimates. Finally, our experiments indicate that the test of overidentifying restrictions is well behaved and can provide a suitable metric of misspecification.

7 Application: Fuhrer and Olivei (2004) revisited

The popular New-Keynesian framework for monetary policy analysis combines mixed backward/forward-looking, micro-founded, output (IS curve) and inflation (Phillips curve) Euler equations with a policy reaction function. This elementary three equation model is the cornerstone of an extensive literature that investigates optimal monetary policy (see Taylor's 1999 edited volume and Walsh's 2003 textbook, chapter 11, and references therein). The stability of alternative policy designs depends crucially on the relative weight of the backward and forward-looking elements and is an issue that has to be determined empirically for central banking is foremost, a practical matter.

However, estimating these relationships empirically is complicated by the poor sample properties of popular estimators. Fuhrer and Olivei (2004) discuss the weak instrument problem that characterizes GMM in this type of application and then propose a GMM variant where the dynamic constraints of the economic model are imposed on the instruments. They dub this procedure "optimal instruments" GMM ($OI - GMM$) and explore its properties relative to conventional GMM and MLE estimators.

We find it is useful to apply PMD to the same examples Fuhrer and Olivei (2004) analyze to provide the reader a context of comparison for our method. The basic specification is (using the same notation as in Fuhrer and Olivei, 2004):

$$z_t = (1 - \mu) z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t \tag{43}$$

In the output Euler equation, z_t is a measure of the output gap, x_t is a measure of the real interest rate, and hence, $\gamma < 0$. In the inflation Euler version of (43), z_t is a measure of inflation, x_t is a measure of the output gap, and $\gamma > 0$ signifying that a positive output gap exerts "demand pressure" on inflation.

Fuhrer and Olivei (2004) experiment with a quarterly sample from 1966:Q1 to 2001:Q4 and use the following measures for z_t and x_t. The output gap is measured, either by the log deviation of real GDP from its Hodrick-Prescott (HP) trend or, from a segmented time trend (ST) with breaks in 1974 and 1995. Real interest rates are measured by the difference of the federal funds rate and next period's inflation. Inflation is measured by the log change in the GDP, chain-weighted price index. In addition, Fuhrer and Olivei (2004) experiment with real unit labor costs (RULC) instead of the output gap for the inflation Euler equation. Further details can be found in their paper.

We begin by recasting expression (43) in terms of the set-up used in earlier sections and to make the connections explicit. Hence, let $\mathbf{y}_t = (z_t \quad x_t)'$ so that

$$\mathbf{y}_t = \Phi_1' \mathbf{y}_{t-1} + \Phi_2' E_t \mathbf{y}_{t+1} + \mathbf{u}_t$$

and in particular,

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^1 & \phi_{21}^1 \\ \phi_{12}^1 & \phi_{22}^1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^2 & \phi_{21}^2 \\ \phi_{12}^2 & \phi_{22}^2 \end{pmatrix} \begin{pmatrix} E_t z_{t+1} \\ E_t x_{t+1} \end{pmatrix} + \begin{pmatrix} u_{zt} \\ u_{x_t} \end{pmatrix}$$

in the notation of expression (1). Notice that expression (43) imposes the following parameter constraints

$$\begin{array}{ll} \phi_{11}^1 = (1-\mu) & \phi_{11}^2 = \mu \\ \phi_{21}^1 = 0 & \phi_{21}^2 = \gamma \end{array}, \tag{44}$$

leaving the parameters of the second equation unconstrained. Although one could estimate the first equation in isolation, as Fuhrer and Olivei (2004) and many others do, we preferred to estimate both equations jointly as a way to improve the quality of our estimates and the chances of passing our specification test since this model is notoriously difficult to fit.

The parameter vector $\phi = vec(\Phi_1 \quad \Phi_2)$ is therefore an 8×1 vector, with the first four elements corresponding to the Euler conditions in expression (43) and the constraints in expression (44), and where the second four elements correspond to the parameters of the expression for x_t, which are not of immediate interest. A simple way to implement the constraints in expression (44) is by defining:

$$C = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; c = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

so that

$$C\phi = c \tag{45}$$

imposes the constraints in (44).

In this univariate example, the contemporaneous parameter restrictions are very simple to incorporate since $P_z = \sigma_{u_z}$ and hence

$$\left(\frac{1}{\sigma_{u_z}} - 1\right) = \phi_{11}^2 b_{11}^1 + \phi_{21}^2 b_{12}^1,$$

where $V\left(\frac{1}{\sigma_{u_z}} - 1\right) = 2$. Theorem 6 in section 4.3 provides the necessary results to estimate the model in (43) by PMD. Using the notation in that section and noticing that the linearity of the problem means we can write

$$\begin{bmatrix} f\left(\widehat{b}_T;\phi\right) \\ g\left(\widehat{b}_T;\phi\right) \end{bmatrix} = H_\phi\left(\widehat{b}_T\right)\phi = \widehat{H}_\phi\phi,$$

then the minimum-distance problem subject to the constraints in (45) is

$$\min_{\phi}\left[S\widehat{b}_T - \widehat{H}_\phi\phi\right]'\widehat{W}_q\left[S\widehat{b}_T - \widehat{H}_\phi\phi\right]$$

$$s.t.$$

$$C\phi = c$$

where

$$\widehat{W}_q = \begin{bmatrix} 1/2 & \mathbf{0} \\ \mathbf{0} & \left(S\Omega_B S'\right)^{-1} \end{bmatrix}.$$

The solution of the Lagrangian of the constrained model is

$$\widehat{\phi}_c = \widehat{\phi}_T - \left(\widehat{H}_\phi' \widehat{W}_q \widehat{H}_\phi\right)^{-1} C' \left[C \left(\widehat{H}_\phi' \widehat{W}_q \widehat{H}_\phi\right)^{-1} C'\right]^{-1} \left(C\widehat{\phi}_T - c\right)$$

where $\widehat{\phi}_c$ denotes the constrained estimate of ϕ, and $\widehat{\phi}_T$ denotes the unconstrained estimate. This result should look very familiar as it is a generalization of the well-known restricted least-squares result. Similarly, it is easy to show that the covariance matrix of the restricted estimates can be calculated as

$$\begin{aligned} \widehat{\Omega}_{\widehat{\phi}_c} &= \Xi \widehat{\Omega}_{\widehat{\phi}} \Xi' \\ \Xi &= \left[I - \left(\widehat{H}_\phi' \widehat{W}_q \widehat{H}_\phi\right)^{-1} C' \left[C \left(\widehat{H}_\phi' \widehat{W}_q \widehat{H}_\phi\right)^{-1} C'\right]^{-1} C\right] \end{aligned}$$

Since the second term in brackets is a positive definite matrix, it is easy to see that the variance of the constrained estimator is smaller than the variance of the unconstrained estimator.

Table 5 and Figure 1 summarize the empirical estimates of the output Euler equation and correspond to the results in Table 4 in Fuhrer and Olivei (2004), where as Tables 6 and 7 and Figure 2 summarize the estimates of the inflation Euler equation and correspond to the results in Table 5 in Fuhrer and Olivei (2004).

For each Euler equation, we report GMM, MLE, $OI - GMM$, estimates that replicate those in Fuhrer and Olivei (2004). Next, we report PMD results based on $h = 20$. Figures 1 and 2 display the estimates of μ and γ in (43) as a function of h and the associated two-standard error bands. Perhaps with the exception of γ in the RULC specification of Figure 2, the graphs show that the parameter estimates vary very little with h even though the standard errors get somewhat narrower. PMD results are reported for the constrained and unconstrained versions of the Euler equation and we also report the overall specification test for the unconstrained model as a function of h so as to stack the odds in favor of the null that the model is correctly specified.

We begin with a general overview of the results. Since the true model is unknowable, there is no definitive metric by which one method can be judged to offer closer estimates to the true parameter values. However PMD estimates do not depart wildly from the estimates reported by the alternative methods. In almost all the cases, we found the overall specification test rejects the proposed Euler specification. PMD results are generally closer to values that would be expected from economic theory (as much as this finding can be of comfort) and unconstrained estimates are generally near the values for the constrained estimates so that the constraints are generally not rejected by the data. Estimates that included the contemporaneous correlation restrictions where virtually identical to estimates that excluded these conditions. Therefore and to make the results more comparable to GMM, we report the results based on excluding these contemporaneous conditions.

PMD estimates for γ in the output Euler equation in Table 5 (-0.15 and -0.20 for the HP and ST specifications respectively) are two orders of magnitude larger than conventional estimates (which are in the range 0.0024 to -0.0084) and of the correct sign. While statistically this coefficient is not significant, the magnitude of the coefficient is economically plausible. The unconstrained version of the parameter estimates suggest that γ may be even larger in magnitude (-0.54 to -0.64) and statistically significant. The unconstrained estimates also suggest that the backward/forward looking terms are approximately of the same magnitude (0.48 vs. 0.45 for HP; 0.42 vs. 0.46 for ST) and they add up to 0.93/0.88 (HP/ST), very close to the canonical value of 1. However, these estimates also suggest a possibly non-zero coefficient on the lagged value of the real interest rate (0.46/0.47 for HP/ST). Unfortunately, the overall specification test strenuously rejects the model, which makes difficult any forcible interpretation of the estimates.

PMD estimates of γ in the inflation Euler equation are very close to those estimated by MLE or $OI - GMM$. In fact, estimates of μ and γ for the RULC model are virtually identical. The unconstrained estimates suggest the ratio of backward/forward looking terms across specifications is approximately 0.45/0.25 and adds up to about 0.70, somewhat further from the canonical value of 1 but within statistical bounds. Unconstrained estimates of the lagged output gap term are close to zero (except for the RULC specification) and the coefficient of γ is estimated to be about 0.10 (but not statistically significant) for the HP and ST specifications and 0.21 (and significant) for the RULC specification. The overall specification test rejects the model except at horizon 4 for all specifications and horizons 7 and 8 for the HP and ST specifications. Parameter estimates at these horizons are very similar to the final estimates reported in Table 6 and hence are not reported separately.

Summarizing, we find PMD provides estimates that are at times similar to estimates by other methods, at times quite different but in directions that would be predicted by economic theory. PMD estimates are more stable across specifications and with respect to the unconstrained versions of the model. The overall specification test rejects the Euler specifications most of the time and although this makes comparisons across methods difficult, PMD appears to perform well. An exhaustive comparative study across methods can only be done with extensive Monte Carlo simulations, but this is beyond the scope of this paper. Despite the apparent inconclusiveness of these results, we wish to point out that PMD has better theoretical properties than the alternative methods considered and PMD would allow further investigation with auxiliary conditioning variables (in the form of a vector $\mathbf{y}_{2t}$ in the notation of previous sections).

8 Conclusions

This paper introduces a disarmingly simple, two-step, minimum-distance method to estimate dynamic systems of equations. The premise of the method is to remain agnostic with respect to the dynamics and the variables that may have been omitted from the candidate model specified with the objective of obtaining consistent parameter estimates nevertheless. The principle behind the method consists in matching the impulse responses of the data estimated semi-parametrically with the impulse responses implied by the candidate model specified – the dimension along which most macroeconomic models are evaluated. Consequently, the method provides a simple chi-square test that measures the distance between the data's and the model's impulse responses and which can be used as an omnibus misspecification test.

An important feature of the method is the first-stage, semi-parametric estimator of the impulse response function. We show that this estimator is consistent and asymptotically normal and derive the analytic covariance matrix of the impulse response coefficients across time and across variables. On its own, we view this as an important contribution to empirical macroeconomic research: not only it allows impulse responses to be estimated without a reference model, it provides simple analytical results to do joint inference.

There are many research questions space constraints prevent us from exploring in this paper and that we open as topics for future research. First, it is natural to extend PMD to nonlinear models. Theorems 4-6 are derived for generic functions relating the impulse responses and the structural parameters of interest and therefore immediately encompass nonlinear specifications. However, depending on the nature of the nonlinearities, it seems natural to extend and estimate the first-stage impulse responses flexibly along the lines in Jordà (2005). Second, it is desirable to derive asymptotic results that offer guidance on the optimal rate at which $h \to \infty$ with the sample size and confirm with Monte Carlo experimentation, an appropriate practical rule-of-thumb. Third, it is important to determine the power properties of the overall specification test in light of the small sample deficiencies of its GMM cousin. The Monte Carlo results that we offer here suggest the test has good properties but a more exhaustive investigation is needed. Fourth, we hope PMD will be applied widely and as more applications are developed, a more comprehensive investigation of the practical merits of PMD relative to MLE and GMM seems warranted. Fifth, PMD appears well suited to estimate $VARMA(p,q)$ models, which are often difficult to estimate because of numerical instabilities when maximizing the likelihood in large systems. Since PMD involves two simple least-squares steps, we expect PMD to offer advantages in this dimension. In addition, we expect that PMD can be extended to other less conventional time-series models, such as multivariate $GARCH$ specifications, that are also difficult to estimate in practice.

9 Appendix

Proof. Theorem 1

Notice that

$$\begin{aligned}\widehat{A}(k,h)-A(k,h) &= \widehat{\Gamma}'_{1,k,h}\widehat{\Gamma}_k^{-1}-A(k,h)\widehat{\Gamma}_k\widehat{\Gamma}_k^{-1}= \\ &\left\{(T-k-h)^{-1}\sum_{j=k}^{\infty}\mathbf{v}_{k,t+h}X'_{t,k}\right\}\widehat{\Gamma}_k^{-1}\end{aligned}$$

where

$$\mathbf{v}_{k,t+h}=\sum_{j=k+1}^{\infty}A_j^h\mathbf{y}_{t-j}+\varepsilon_{t+h}+\sum_{j=1}^{h-1}B_j\varepsilon_{t+h-j}$$

Hence,

$$\begin{aligned}\widehat{A}(k,h)-A(k,h) &= \left\{(T-k-h^{-1})\sum_{t=k}^{T-h}\left(\sum_{j=k+1}^{\infty}A_j^h\mathbf{y}_{t-j}\right)X'_{t,k}\right\}\widehat{\Gamma}_k^{-1}+ \\ &\left\{(T-k-h^{-1})\sum_{t=k}^{T-h}\varepsilon_{t+h}X'_{t,k}\right\}\widehat{\Gamma}_k^{-1}+ \\ &\left\{(T-k-h^{-1})\sum_{t=k}^{T-h}\left(\sum_{j=1}^{h}B_j\boldsymbol{\varepsilon}_{t+h-j}\right)X'_{t,k}\right\}\widehat{\Gamma}_k^{-1}\end{aligned}$$

Define the matrix norm $\|C\|_1^2=\sup_{l\neq 0}\frac{l'C'C'}{l'l}$, that is, the largest eigenvalue of $C'C$. When C is symmetric, this is the square of the largest eigenvalue of C. Then

$$\|AB\|^2\leq\|A\|_1^2\|B\|^2 \text{ and } \|AB\|^2\leq\|A\|^2\|B\|_1^2$$

Hence

$$\left\|\widehat{A}(k,h)-A(k,h)\right\|\leq\|U_{1T}\|\left\|\widehat{\Gamma}_k^{-1}\right\|_1+\|U_{2T}\|\left\|\widehat{\Gamma}_k^{-1}\right\|_1+\|U_{3T}\|\left\|\widehat{\Gamma}_k^{-1}\right\|_1$$

where

$$\begin{aligned}U_{1T} &= \left\{(T-k-h^{-1})\sum_{t=k}^{T-h}\left(\sum_{j=k+1}^{\infty}A_j^h\mathbf{y}_{t-j}\right)X'_{t,k}\right\} \\ U_{2T} &= \left\{(T-k-h^{-1})\sum_{t=k}^{T-h}\varepsilon_{t+h}X'_{t,k}\right\} \\ U_{3T} &= \left\{(T-k-h^{-1})\sum_{t=k}^{T-h}\left(\sum_{j=1}^{h}B_j\boldsymbol{\varepsilon}_{t+h-j}\right)X'_{t,k}\right\}\end{aligned}$$

Lewis and Reinsel (1985) show that $\left\|\widehat{\Gamma}_k^{-1}\right\|_1$ is bounded, therefore, the next objective is to show $\|U_{1T}\| \xrightarrow{p} 0$, $\|U_{2T}\| \xrightarrow{p} 0$, and $\|U_{3T}\| \xrightarrow{p} 0$. We begin by showing $\|U_{2T}\| \xrightarrow{p} 0$, which is easiest to see since ε_{t+h} and $X'_{t,k}$ are independent, so that their covariance is zero. Formally and following similar derivations in Lewis and Reinsel (1985)

$$E\left(\|U_{2T}\|^2\right) = (T-k-h)^{-2} \sum_{t=k}^{T-h} E\left(\varepsilon_{t+h}\varepsilon'_{t+h}\right) E(X'_{t,k}X'_{t,k})$$

by independence. Hence

$$E\left(\|U_{2T}\|^2\right) = (T-k-h)^{-1} tr(\Sigma) k \left\{tr\left[\Gamma(0)\right]\right\}$$

Since $\frac{k}{T-k-h} \rightarrow 0$ by assumption (ii), then $E\left(\|U_{2T}\|^2\right) \xrightarrow{p} 0$, and hence $\|U_{2T}\| \xrightarrow{p} 0$.

Next, consider $\|U_{3T}\| \xrightarrow{p} 0$. The proof is very similar since $\varepsilon_{t+h-j},\ j = 1, ..., h-1$ and $X'_{t,k}$ are independent. As long as $\|B_j\|^2 < \infty$ (which is true given that the Wold decomposition ensures that $\sum_{j=0}^{\infty} \|B_j\| < \infty$, then using the same arguments we used to show $\|U_{2T}\| \xrightarrow{p} 0$, it is easy to see that $\|U_{3T}\| \xrightarrow{p} 0$.

Finally, we show that $\|U_{1T}\| \xrightarrow{p} 0$. The objective here is to show that assumption (iii) implies that

$$k^{1/2} \sum_{j=k+1}^{\infty} \left\|A_j^h\right\| \rightarrow 0,\ \ k, T \rightarrow 0$$

because we will need this condition to hold to complete the proof later. Recall that

$$A_j^h = B_{h-1}A_j + A_{j+1}^{h-1};\ \ A_{j+1}^0 = 0;\ \ B_0 = I_r;\ \ h, j \geq 1,\ \ h \text{ finite}$$

Hence

$$k^{1/2} \sum_{j=k+1}^{\infty} \left\|A_j^h\right\| = k^{1/2} \left\{ \sum_{j=k+1}^{\infty} \left\|B_{h-1}A_j + B_{h-2}A_{j+1} + ... + B_1 A_{j+h-2} + A_{j+h-1}\right\| \right\}$$

by recursive substitution. Thus

$$k^{1/2} \sum_{j=k+1}^{\infty} \left\|A_j^h\right\| \leq k^{1/2} \left\{ \sum_{j=k+1}^{\infty} \left\|B_{h-1}A_j\right\| + ... + \left\|B_1 A_{j+h-2}\right\| + \left\|A_{j+h-1}\right\| \right\}$$

Define λ as the $\max\{\|B_{h-1}\|, ..., \|B_1\|\}$, then since $\sum_{j=0}^{\infty}\|B_j\| < \infty$ we know $\lambda < \infty$ so that

$$k^{1/2}\sum_{j=k+1}^{\infty}\left\|A_j^h\right\| \leq k^{1/2}\left\{\lambda\sum_{j=k+1}^{\infty}\|A_j\| + ... + \lambda\sum_{j=k+1}^{\infty}\|A_{j+h-2}\| + \sum_{j=k+1}^{\infty}\|A_{j+h-1}\|\right\}$$

By assumption (iii) and since $\lambda < \infty$, then each of the elements in the sum goes to zero as T, k go to infinity. Finally, to prove $\|U_{1T}\| \xrightarrow{p} 0$ all that is required is to follow the same steps as in Lewis and Reinsel (1985) but using the condition

$$k^{1/2}\sum_{j=k+1}^{\infty}\left\|A_j^h\right\| \to 0, \ k, T \to 0$$

instead. ■

Proof. Theorem 2

We begin by showing that $W_{1T} \xrightarrow{p} 0$. Lewis and Reinsel (1985) show that under assumption (ii), $k^{1/2}\left\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \xrightarrow{p} 0$ and $E\left(\left\|k^{-1/2}(T-k-h)^{1/2}U_{1T}\right\|\right) \leq s(T-k-h)^{1/2}\sum_{j=k+1}^{\infty}\left\|A_j^h\right\| \xrightarrow{p} 0$; $k, T \to \infty$ from assumption (iii) and using similar derivations as in the proof of consistency with s being a generic constant. Hence $W_{1T} \xrightarrow{p} 0$.

Next, we show $W_{2T} \xrightarrow{p} 0$. Notice that

$$|W_{2T}| \leq k^{1/2}\left\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \left\|k^{-1/2}(T-k-h)^{1/2}U_{2T}^*\right\|$$

As in the previous step, Lewis and Reinsel (1985) establish that $k^{1/2}\left\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \xrightarrow{p} 0$ and from the proof of consistency, we know the second term is bounded in probability, which is all we need to establish the result.

Lastly, we need to show $W_{3T} \xrightarrow{p} 0$, however, the proof of this result is identical to that in Lewis and Reinsel once one realizes that assumption (iii) implies that

$$(T-k-h)^{1/2}\sum_{j=k+1}^{\infty}\left\|A_j^h\right\| \xrightarrow{p} 0$$

and substituting this result into their proof. ■

Proof. Theorem 3

Follows directly from Lewis and Reinsel (1985) by redefining

$$A_{T_m} = (T-k-h)^{1/2}\, vec\left[\left\{(T-k-h)^{-1}\sum_{t=k}^{T-h}\left(\varepsilon_{t+h}+\sum_{j=1}^{h-1}B_j\varepsilon_{t+h-j}\right)X'_{t,k}(m)\right\}\Gamma_k^{-1}\right]$$

for $m = 1, 2, ...$ and $X_{t,k}(m)$ as defined in Lewis and Reinsel (1985). ■

Proof. Theorem 4

Recall

$$S\widehat{b}_T \equiv (R\otimes S_0)\,\widehat{b}_T$$

and for notational convenience, define

$$g_1\left(\widehat{b}_T\right) = (I_r\otimes(S_1\mathbf{B}R \quad S_2\mathbf{B}R))$$

Since $\widehat{b}_T \xrightarrow{p} b_0$, then

$$S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\phi \xrightarrow{p} Sb_0 - g_1(b_0)\,\phi$$

by the continuous mapping theorem. Furthermore and given assumption (i)

$$\begin{aligned}\widehat{Q}_T(\phi) &= \left[S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\phi\right]'\widehat{W}\left[S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\phi\right] \xrightarrow{p} \\ [Sb_0 - g_1(b_0)\,\phi]'\,W\,[Sb_0 - g_1(b_0)\,\phi] &\equiv Q_0(\phi)\end{aligned}$$

which is a quadratic expression that is maximized at ϕ_0. Assumption (v) provides a necessary condition for identification of the parameters (i.e., that there be at least as many moment matching conditions as parameters) that must be satisfied to establish uniqueness. As a quadratic function, $Q_0(\phi)$ is obviously a continuous function. The last thing to show is that

$$\sup_{\phi\in\Theta}\left|\widehat{Q}_T(\phi) - Q_0(\phi)\right| \xrightarrow{p} 0$$

uniformly.

For compact Θ and continuous $Q_0(\phi)$, Lemma 2.8 in Newey and McFadden (1994) provides that this condition holds if and only if $\widehat{Q}_T(\phi) \xrightarrow{p} Q_0(\phi)$ for all ϕ in Θ and $\widehat{Q}_T(\phi)$ is stochastically equicontinuous. The former has already been established, so it remains to show stochastic equicontinuity of $\widehat{Q}_T(\phi)$.[3] Notice that

$$\left|\widehat{Q}_T(\tilde{\phi}) - \widehat{Q}_T(\phi)\right| = \left| \begin{array}{c} \left[S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\tilde{\phi}\right]' \widehat{W} \left[S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\tilde{\phi}\right] - \\ \left[S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\phi\right]' \widehat{W} \left[S\widehat{b}_T - g_1\left(\widehat{b}_T\right)\phi\right] \end{array} \right|$$

$$= \left|(\tilde{\phi}-\phi)' g_1(\widehat{b}_T)'\widehat{W} g_1(\widehat{b}_T)(\tilde{\phi}-\phi) + (S\widehat{b}_T - g_1(\widehat{b}_T)\phi)'(\widehat{W}+\widehat{W}')g_1(\widehat{b}_T)(\tilde{\phi}-\phi)\right|$$

$$\leq \left|\left|g_1(\widehat{b}_T)'\widehat{W} g_1(\widehat{b}_T)\right|\right| \; \left|\left|\tilde{\phi}-\phi\right|\right|^2 + \left|\left|(S\widehat{b}_T - g_1(\widehat{b}_T)\phi)'(\widehat{W}+\widehat{W}')g_1(\widehat{b}_T)\right|\right| \; \left|\left|\tilde{\phi}-\phi\right|\right|.$$

Define

$$\widehat{R}_T \equiv \left|\left|g_1(\widehat{b}_T)'\widehat{W} g_1(\widehat{b}_T)\right|\right| + \left|\left|(S\widehat{b}_T - g_1(\widehat{b}_T)\phi)'(\widehat{W}+\widehat{W}')\right|\right|$$

and

$$\alpha = \underset{\delta \in \{1,2\}}{argmax} \; \{\left|\left|\tilde{\phi}-\phi\right|\right|^{\delta}\}.$$

Given $\widehat{b}_T \xrightarrow{p} b_0$ and (i) and since $||.||$ is a continuous function, $\widehat{R}_T = O_p(1)$ and there exists an M such that Prob$(\widehat{R}_T > M) < \eta$ for all n large enough. Let

$$\begin{aligned} \widehat{\Delta}_T &= \epsilon \widehat{R}_T / M \\ N &= \{\tilde{\phi} : \left|\left|\tilde{\phi}-\phi\right|\right|^{\alpha} \leq \epsilon/M\}. \end{aligned}$$

Then, $Prob(|\widehat{\Delta}_T| > \epsilon) = Prob(|\widehat{R}_T| > M) < \eta$ and for all $\tilde{\phi}, \phi \in N$, $|\widehat{Q}_T(\tilde{\phi}) - \widehat{Q}_T(\theta)| \leq \widehat{R}_T ||\tilde{\phi} - \phi||^{\alpha} \leq \widehat{\Delta}_T$. ■

Proof. Theorem 5

Under assumption (iii) b_0 and ϕ_0 are in the interior of their parameter spaces and by assumption (ii) $\widehat{b}_T \xrightarrow{p} b_0$, $\widehat{\phi}_T \xrightarrow{p} \phi_0$. Further, by assumption (iv), $g(\widehat{b}_T;\phi)$ is continuously differentiable in a

[3] Stochastic equicontinuity: For every $\epsilon, \eta > 0$ there exists a sequence of random variables $\hat{\Delta}_t$ and a sample size t_0 such that for $t \geq t_0$, Prob$(|\hat{\Delta}_T| > \epsilon) < \eta$ and for each ϕ there is an open set N containing ϕ with $\sup_{\tilde{\phi} \in N} \left|\widehat{Q}_T(\tilde{\phi}) - \widehat{Q}_T(\phi)\right| \leq \hat{\Delta}_T$, for $t \geq t_0$.

neighborhood of b_0 and ϕ_0 and hence $\widehat{\phi}_T$ solves the first order conditions of the minimum-distance problem

$$\min_{\phi}\left[S\widehat{b}_T-g(\widehat{b}_T;\phi)\right]'\widehat{W}\left[S\widehat{b}_T-g(\widehat{b}_T;\phi)\right]$$

which are

$$-G_{\phi}\left(\widehat{b}_T;\widehat{\phi}_T\right)'\widehat{W}\left[S\widehat{b}_T-g(\widehat{b}_T;\widehat{\phi}_T)\right]=0$$

By assumption (iv), these first order conditions can be expanded about ϕ_0 in mean value expansion

$$g(\widehat{b}_T;\widehat{\phi}_T)=g(\widehat{b}_T;\phi_0)+G_{\phi}\left(\widehat{b}_T;\overline{\phi}\right)\left(\widehat{\phi}_T-\phi_0\right)$$

where $\overline{\phi}\in[\widehat{\phi}_T,\phi_0]$. Similarly, a mean value expansion of $g(\widehat{b}_T;\phi_0)$ around b_0 is

$$g(\widehat{b}_T;\phi_0)=g(b_0;\phi_0)+G_b\left(\overline{b};\phi_0\right)\left(\widehat{b}_T-b_0\right)$$

Combining both mean value expansions and multiplying by $\sqrt{T}$, we have

$$\begin{aligned}\sqrt{T}g(\widehat{b}_T;\widehat{\phi}_T) &= \sqrt{T}g(b_0;\phi_0)+G_{\phi}\left(\widehat{b}_T;\overline{\phi}\right)\sqrt{T}\left(\widehat{\phi}_T-\phi_0\right)+\\ &\quad G_b\left(\overline{b};\phi_0\right)\sqrt{T}\left(\widehat{b}_T-b_0\right)\end{aligned}$$

Since $\overline{b}\in[\widehat{b}_T,b_0]$, $\overline{\phi}\in[\widehat{\phi}_T,\phi_0]$ and $\widehat{b}_T\xrightarrow{p}b_0$, $\widehat{\phi}_T\xrightarrow{p}\phi_0$ then, along with assumption (iv), we have

$$\begin{aligned}G_{\phi}\left(\widehat{b}_T;\overline{\phi}\right)\xrightarrow{p}G_{\phi}\left(b_0;\phi_0\right) &= G_{\phi}\\ G_b\left(\overline{b};\phi_0\right)\xrightarrow{p}G_b(b_0;\phi_0) &= G_b\end{aligned}$$

and hence

$$\sqrt{T}g(\widehat{b}_T;\widehat{\phi}_T)=\sqrt{T}g(b_0;\phi_0)+G_{\phi}\sqrt{T}\left(\widehat{\phi}_T-\phi_0\right)+G_b\sqrt{T}\left(\widehat{b}_T-b_0\right)+o_p(1)$$

In addition, by assumption (i) $\widehat{W}\xrightarrow{p}W$ and notice that $g\left(b_0,\phi_0\right)=Sb_0$, which combined with the first order conditions and the mean value expansions described above, allow us to write

$$-G_{\phi}'W\left[\sqrt{T}\left(S\widehat{b}_T-\left\{Sb_0+G_{\phi}\sqrt{T}\left(\widehat{\phi}_T-\phi_0\right)+G_b\sqrt{T}\left(\widehat{b}_T-b_0\right)\right\}\right)\right]=o_p(1)$$

Since we know that

$$\sqrt{T}\left(\widehat{b}_T - b_0\right) \xrightarrow{d} N\left(0, \Omega_B\right)$$

then

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} \left(G_\phi' W G_\phi\right)^{-1}\left\{G_\phi' W S + G_\phi' W G_b\right\}\sqrt{T}\left(\widehat{b}_T - b_0\right)$$

by assumption (vii) which ensures that $G_\phi' W G_\phi$ is invertible and assumption (viii) ensures identification. Therefore, from the previous expression we arrive at

$$\begin{aligned}
& \sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} N\left(0, \Omega_\phi\right) \\
\Omega_\phi \quad = \quad & \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W S \Omega_B S' W G_\phi\right)\left(G_\phi' W G_\phi\right)^{-1} + \\
& \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W G_b \Omega_B G_b' W G_\phi\right)\left(G_\phi' W G_\phi\right)^{-1} - \\
& \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W S \Omega_B G_b' W G_\phi\right)\left(G_\phi' W G_\phi\right)^{-1} - \\
& \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W G_b \Omega_B S' W G_\phi\right)\left(G_\phi' W G_\phi\right)^{-1}
\end{aligned}$$

By assumption (i), we choose the optimal weighting matrix $W = \left(S \Omega_B S'\right)^{-1}$ and hence the variance of $\widehat{\phi}_T$ simplifies to the final expression in the theorem, that is

$$\begin{aligned}
\Omega_\phi \quad = \quad & \left(G_\phi' W G_\phi\right)^{-1} + \\
& \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W G_b \Omega_B G_b' W G_\phi\right)\left(G_\phi' W G_\phi\right)^{-1} - \\
& \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W S \Omega_B G_b' W G_\phi\right)\left(G_\phi' W G_\phi\right)^{-1} - \\
& \left(G_\phi' W G_\phi\right)^{-1}\left(G_\phi' W G_b \Omega_B S' W G_\phi\right)\left(G_\phi' W G_\phi\right)
\end{aligned}$$

■

Proof. Theorem 6

We begin by deriving the distribution of $\widehat{Q} = \left(R' \widehat{P} R\right)^{-1} - I$ where $\widehat{P}\widehat{P}' = \widehat{\Sigma}_\varepsilon$. The distribution of $vech(\widehat{\Sigma}_\varepsilon)$ is directly available in Lütkepohl (1993) and is

$$\begin{aligned}
& \sqrt{T}\left(vech\left(\widehat{\Sigma}_\varepsilon\right) - vech\left(\Sigma_\varepsilon\right)\right) \xrightarrow{d} N\left(0, \Omega_\sigma\right) \\
\Omega_\sigma \quad = \quad & 2 D_r^+ \left(\Sigma_\varepsilon \otimes \Sigma_\varepsilon\right) D_r^{+'}
\end{aligned}$$

where $D_r^+ = (D_r' D_r)^{-1} D_r$ and D_r is the duplication matrix such that for any square, $r \times r$ matrix Σ, then $vec(\Sigma) = D_r vech(\Sigma)$. In addition, Lütkepohl (1993) also shows that

$$\frac{\partial vec(P)}{\partial vech(\Sigma_\varepsilon)} = L_r' \left\{ L_r \left(I_{r^2} + K_{rr} \right) \left(P \otimes I_r \right) L_r' \right\}^{-1}$$

where L_r is the elimination matrix such that for any square, $r \times r$ matrix Σ then $vech(\Sigma) = L_r vec(\Sigma)$ and K_{rr} is the commutation matrix such that $vec(\Sigma') = K_{rr} vec(\Sigma)$. All that remains therefore is to derive $\partial vec(Q) / \partial vec(P)$. Notice that

$$dQ = -\left(R'PR\right)^{-1} R'dPR \left(R'PR\right)^{-1}$$

and hence

$$\begin{aligned} dvecQ &= -[\left(RP'R'\right)^{-1} \otimes \left(R'PR\right)^{-1}] dvec\left(R'PR\right) \\ dvec(R'PR) &= \left[R' \otimes R\right] dvec(P) \end{aligned}$$

Combining terms

$$dvec(Q) = -[\left(RP'R'\right)^{-1} \otimes \left(R'PR\right)^{-1}] \left[R' \otimes R\right] dvec(P)$$

which allows us to arrive at the final result that

$$\begin{aligned} \Omega_q &= 2\Gamma D_r^+ \left(\Sigma_\varepsilon \otimes \Sigma_\varepsilon\right) D_r^{+'} \Gamma' \\ \Gamma &= [\left(RP'R'\right)^{-1} \otimes \left(R'PR\right)^{-1}] \left[R' \otimes R\right] L_r' \left\{ L_r \left(I_{r^2} + K_{rr} \right) \left(P \otimes I_r \right) L_r' \right\}^{-1} \end{aligned}$$

Lütkepohl (1993) also shows that

$$\sqrt{T} \begin{pmatrix} vec(\widehat{B}_1) - vec(B_1) \\ vech(\widehat{\Sigma}_\varepsilon) - vech(\Sigma_\varepsilon) \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{B_1} & 0 \\ 0 & \Omega_\sigma \end{pmatrix} \right)$$

from where it is easy to see the justification for assumption (v) that the covariance of $\widehat{q}_T$ and $\widehat{b}_T$ is zero. With these results established, the proof of theorem 6 proceeds along the same lines as the proof of theorem 5, that is, under the assumptions of theorem 6, $\widehat{\phi}_T$ will be a solution to the minimum-distance problem expanded to include the contemporaneous correlations. Then we take a mean value expansion of the first order conditions and given that W_q is the optimal weighting matrix, it is straightforward to derive the desired result. ■

References

Anderson, Theodore W. (1994) **The Statistical Analysis of Time Series Data**. New York, New York: Wiley Interscience.

Christiano, Lawrence J. (2002) "Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients, *Computational Economics*, 20, 21-55.

Evans, George W. and Seppo Honkapohja (2001) **Learning and Expectations in Macroeconomics.** Princeton and Oxford: Princeton University Press.

Farmer, Roger E. A. (1993) **The Macroeconomics of Self-Fulfilling Prophecies.** Cambridge, MA: MIT Press.

Ferguson, Thomas S. (1958) "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics*, 29: 1046-62.

Fuhrer, Jeffrey C. and Giovanni P. Olivei (2004) "Estimating Forward Looking Euler Equations with GMM Estimators: An Optimal Instruments Approach," FRB Boston Series, paper n. 04-2.

Hamilton, James D. (1994) **Time Series Analysis**. Princeton, New Jersey: Princeton University Press.

Hurvich, Clifford M. and Chih-Ling Tsai (1989) "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76(2): 297-307.

Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1): 161-182.

Lewis, R. A. and Gregory C. Reinsel (1985) "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16(33): 393-411.

Lütkepohl, Helmut (1993) **Introduction to Multiple Time Series Analysis**, (2nd ed.). Berlin: Springer-Verlag.

Newey, Whitney K. and Daniel L. McFadden (1994) "Large Sample Estimation and Hypothesis Testing," in **Handbook of Econometrics, v. 4**, Robert F. Engle and Daniel L. McFadden, (eds.). Amsterdam: North Holland.

Taylor, John B. (ed.) (1999) **Monetary Policy Rules**. Chicago, Illinois: University of Chicago Press for NBER.

Walsh, Carl E. (2003) **Monetary Theory and Policy, second edition.** Cambridge, Massachusetts: The MIT Press.

TABLE 1 – MONTE CARLO EXPERIMENTS: CASE (i)

$\pi_1 = 0.25$ $\theta_1 = 0.5$ **T = 50**

		h = 2		*h = 5*		*h = 10*	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***0.182***	***0.514***	***0.258***	***0.441***	***0.284***	***0.416***
	SE	0.223	0.206	0.214	0.197	0.216	0.190
	SE (MC)	0.303	0.263	0.229	0.203	0.213	0.210
	χ^2-corr	-		0.487		0.534	
	χ^2-incorr	-		0.118		0.168	
MLE	**Est.**	***0.209***	***0.537***	***0.225***	***0.528***	***0.212***	***0.525***
	SE	0.206	0.176	0.205	0.185	0.204	0.185
	SE (MC)	0.292	0.263	0.301	0.256	0.285	0.252

T = 100

		h = 2		*h = 5*		*h = 10*	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***0.226***	***0.503***	***0.260***	***0.475***	***0.245***	***0.465***
	SE	0.151	0.140	0.149	0.134	0.153	0.136
	SE (MC)	0.177	0.162	0.153	0.141	0.151	0.143
	χ^2-corr	-		0.494		0.531	
	χ^2-incorr	-		0.020		0.034	
MLE	**Est.**	***0.237***	***0.510***	***0.248***	***0.503***	***0.237***	***0.502***
	SE	0.143	0.127	0.143	0.128	0.146	0.131
	SE (MC)	0.152	0.138	0.156	0.145	0.148	0.139

T = 400

		h = 2		*h = 5*		*h = 10*	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***0.244***	***0.504***	***0.243***	***0.502***	***0.248***	***0.503***
	SE	0.075	0.069	0.075	0.066	0.075	0.066
	SE (MC)	0.081	0.074	0.072	0.063	0.080	0.073
	χ^2-corr	-		0.507		0.497	
	χ^2-incorr	-		0.000		0.000	
MLE	**Est.**	***0.248***	***0.508***	***0.241***	***0.508***	***0.249***	***0.503***
	SE	0.072	0.064	0.073	0.064	0.072	0.064
	SE (MC)	0.078	0.071	0.071	0.064	0.077	0.071

Notes: 500 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AIC_c, SE refers to the standard error calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard error based on the 500 estimates of the parameter. χ^2-corr. is the Monte Carlo average p-value of the overall misspecification test when the model is correctly specified. χ^2-incorr. is the Monte Carlo average p-value of the overall misspecification test when the model generated is $y_t = 0.25y_{t-1} + 0.50y_{t-2} + u_t$. Notice that for $h = 2$ the model is exactly identified and hence the value of the test is exactly 0. 500 burn-in observations disregarded when generating the data.

TABLE 2 – MONTE CARLO EXPERIMENTS: CASE (ii)

$\pi_1 = 0.5$ $\theta_1 = 0.25$ **T = 50**

		h = 2		h = 5		h = 10	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***0.469***	***0.208***	***0.463***	***0.221***	***0.473***	***0.179***
	SE	0.190	0.200	0.200	0.205	0.205	0.201
	SE (MC)	0.239	0.224	0.203	0.191	0.195	0.208
	χ^2-corr	-		0.549		0.622	
	χ^2-incorr	-		0.437		0.462	
MLE	**Est.**	***0.468***	***0.280***	***0.453***	***0.288***	***0.449***	***0.272***
	SE	0.192	0.200	0.195	0.206	0.195	0.206
	SE (MC)	0.203	0.212	0.207	0.207	0.201	0.223

T = 100

		h = 2		h = 5		h = 10	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***0.488***	***0.255***	***0.494***	***0.233***	***0.479***	***0.263***
	SE	0.133	0.143	0.143	0.149	0.144	0.145
	SE (MC)	0.148	0.159	0.145	0.157	0.142	0.155
	χ^2-corr	-		0.544		0.569	
	χ^2-incorr	-		0.301		0.316	
MLE	**Est.**	***0.484***	***0.272***	***0.488***	***0.274***	***0.465***	***0.269***
	SE	0.132	0.143	0.134	0.145	0.133	0.144
	SE (MC)	0.133	0.149	0.134	0.148	0.128	0.139

T = 400

		h = 2		h = 5		h = 10	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***0.498***	***0.251***	***0.490***	***0.251***	***0.483***	***0.263***
	SE	0.069	0.074	0.075	0.078	0.075	0.077
	SE (MC)	0.072	0.076	0.078	0.076	0.076	0.077
	χ^2-corr	-		0.490		0.452	
	χ^2-incorr	-		0.163		0.130	
MLE	**Est.**	***0.498***	***0.252***	***0.488***	***0.258***	***0.494***	***0.257***
	SE	0.065	0.073	0.066	0.073	0.066	0.072
	SE (MC)	0.067	0.072	0.068	0.071	0.066	0.073

Notes: 500 Monte Carlo replications, 1^{st}-stage regression lag length chosen automatically by AIC_c, SE refers to the standard errors calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard errors based on the 500 estimates of the parameter. χ^2-corr. is the Monte Carlo average p-value of the overall misspecification test when the model is correctly specified. χ^2-incorr. is the Monte Carlo average p-value of the overall misspecification test when the model generated is $y_t = 0.5 y_{t-1} + 0.25 y_{t-2} + u_t$. Notice that for $h = 2$ the model is exactly identified and therefore the test is exactly 0. 500 burn-in observations disregarded when generating the data.

TABLE 3 – MONTE CARLO EXPERIMENTS: CASE (iii)

$\pi_1 = 0$ ***$\theta_1 = 0.5$*** **T = 50**

		h = 2		*h* = 5		*h* = 10	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***-0.072***	***0.561***	***0.056***	***0.393***	***0.120***	***0.312***
	SE	0.355	0.319	0.283	0.260	0.267	0.275
	SE (MC)	0.858	0.801	0.279	0.265	0.229	0.238
	χ^2-corr	-		0.453		0.538	
	χ^2-incorr	-		0.099		0.186	
MLE	**Est.**	**-**	***0.481***	**-**	***0.478***	**-**	***0.487***
	SE	-	0.126	-	0.126	-	0.125
	SE (MC)	-	0.143	-	0.154	-	0.138

T = 100

		h = 2		*h* = 5		*h* = 10	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***-0.044***	***0.511***	***0.024***	***0.459***	***0.046***	***0.441***
	SE	0.235	0.213	0.201	0.176	0.192	0.171
	SE (MC)	0.292	0.262	0.194	0.182	0.188	0.189
	χ^2-corr	-		0.461		0.513	
	χ^2-incorr	-		0.004		0.016	
MLE	**Est.**	**-**	***0.483***	**-**	***0.490***	**-**	***0.497***
	SE	-	0.088	-	0.088	-	0.088
	SE (MC)	-	0.089	-	0.087	-	0.091

T = 400

		h = 2		*h* = 5		*h* = 10	
		π_1	θ_1	π_1	θ_1	π_1	θ_1
PMD	**Est.**	***-0.008***	***0.507***	***-0.005***	***0.503***	***0.003***	***0.488***
	SE	0.113	0.102	0.100	0.086	0.099	0.087
	SE (MC)	0.117	0.105	0.100	0.087	0.099	0.090
	χ^2-corr	-		0.490		0.503	
	χ^2-incorr	-		0.000		0.000	
MLE	**Est.**	**-**	***0.497***	**-**	***0.501***	**-**	***0.496***
	SE	-	0.043	-	0.043	-	0.044
	SE (MC)	-	0.044	-	0.045	-	0.045

Notes: 500 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AIC_c, SE refers to the standard errors calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard errors based on the 500 estimates of the parameter. MLE estimates for the ARMA(1,1) specification failed to converge. Hence we report estimates based on an ARMA(0,1) specification. χ^2-corr. is the Monte Carlo average p-value of the overall misspecification test when the model is correctly specified. χ^2-incorr. is the Monte Carlo average p-value of the overall misspecification test when the model generated is $y_t = 0.25 y_{t-1} + 0.50 y_{t-2} + u_t$. Notice that for $h = 2$ the model is exactly identified and hence the value of the test is exactly 0. 500 burn-in observations disregarded when generating the data.

TABLE 4 – MONTE CARLO EXPERIMENTS: CASE (iv)

π_1 = 0.5 ***Θ_1 = 0***

T = 50

		h = 2		***h = 5***		***h = 10***	
		π_1	***θ_1***	***π_1***	***θ_1***	***π_1***	***θ_1***
PMD	**Est.**	***0.424***	***0.069***	***0.420***	***0.052***	***0.412***	***0.045***
	SE	0.284	0.299	0.265	0.281	0.254	0.257
	SE (MC)	0.432	0.423	0.261	0.245	0.237	0.231
	χ^2-corr	-		0.537		0.594	
	χ^2-incorr	-		0.442		0.475	
MLE	**Est.**	***0.466***	***-***	***0.456***	***-***	***456***	***-***
	SE	0.126	-	0.126	-	0.126	-
	SE (MC)	0.125	-	0.129	-	0.130	-

T = 100

		h = 2		***h = 5***		***h = 10***	
		π_1	***θ_1***	***π_1***	***θ_1***	***π_1***	***θ_1***
PMD	**Est.**	***0.482***	***0.009***	***0.461***	***0.040***	***0.465***	***0.011***
	SE	0.192	0.208	0.184	0.199	0.182	0.192
	SE (MC)	0.217	0.222	0.177	0.176	0.173	0.172
	χ^2-corr	-		0.550		0.562	
	χ^2-incorr	-		0.345		0.330	
MLE	**Est.**	***0.476***	***-0.009***	***0.461***	***-0.033***	***0.477***	***-0.016***
	SE	0.178	0.199	0.181	0.294	0.181	0.201
	SE (MC)	0.193	0.212	0.184	0.211	0.192	0.203

T = 400

		h = 2		***h = 5***		***h = 10***	
		π_1	***θ_1***	***π_1***	***θ_1***	***π_1***	***θ_1***
PMD	**Est.**	***0.490***	***0.012***	***0.494***	***0.009***	***0.478***	***0.011***
	SE	0.091	0.101	0.091	0.100	0.090	0.099
	SE (MC)	0.100	0.103	0.093	0.097	0.092	0.092
	χ^2-corr	-		0.543		0.616	
	χ^2-incorr	-		0.163		0.121	
MLE	**Est.**	***0.490***	***-0.011***	***0.493***	***-0.011***	***0.488***	***-0.014***
	SE	0.088	0.100	0.087	0.099	0.088	0.100
	SE (MC)	0.093	0.104	0.087	0.102	0.084	0.096

Notes: 500 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AIC_c, SE refers to the standard errors calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard errors based on the 500 estimates of the parameter. For T = 50, MLE estimates for the ARMA(1,1) specification failed to converge. We report instead ARMA(1,0) estimates. χ^2-corr. is the Monte Carlo average p-value of the overall misspecification test when the model is correctly specified. χ^2-incorr. is the Monte Carlo average p-value of the overall misspecification test when the model generated is $y_t = 0.5y_{t-1} + 0.25y_{t-2} + u_t$. Notice that for h = 2 the model is exactly identified and hence the value of the test is exactly 0. 500 burn-in observations disregarded when generating the data.

Table 5 – PMD, MLE, GMM and Optimal Instruments GMM: A Comparison

Estimates of Output Euler Equation: 1966:Q1 to 2001:Q4

$$z_t = (1-\mu)z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t$$

Method	Specification	μ (S.E.)	γ (S.E.)
GMM	**HP**	0.52 (0.053)	0.0024 (0.0094)
GMM	**ST**	0.51 (0.049)	0.0029 (0.0093)
MLE	**HP**	0.47 (0.035)	-0.0056 (0.0037)
MLE	**ST**	0.42 (0.052)	-0.0084 (0.0055)
OI-GMM	**HP**	0.47 (0.062)	-0.0010 (0.023)
OI-GMM	**ST**	0.41 (0.064)	-0.0010 (0.022)
PMD (*h = 20*)	**HP**	0.54 (0.11)	-0.15 (0.23)
PMD (*h = 20*)	**ST**	0.54 (0.11)	-0.20 (0.21)

Unconstrained PMD (*h = 20*)

Coefficient	HP Estimate (S.E.)	ST Estimate (S.E.)
z_{t-1}	0.48 (0.15)	0.42 (0.15)
x_{t-1}	0.46 (0.28)	0.47 (0.27)
$E_t z_{t+1}$	0.45 (0.12)	0.46 (0.12)
$E_t x_{t+1}$	-0.54 (0.36)	-0.64 (0.34)

Overall Specification Test by Impulse Response Horizon

HP		ST	
Horizon	**p-value**	**Horizon**	**p-value**
4	0.000	**4**	0.000
5	0.001	**5**	0.000
6	0.001	**6**	0.001
7-20	0.000	**7-20**	0.000

Notes: z_t is a measure of the output gap, x_t is a measure of the real interest rate, and hence economic theory would predict $\gamma < 0$. GMM, MLE, and OI-GMM estimates correspond to estimates reported in Table 4 in Fuhrer and Olivei (2004). PMD estimates reported here are with impulse response horizon $h = 20$. HP refers to Hodrick-Prescott filtered log of real GDP, and ST refers to log of real GDP detrended by a deterministic segmented trend. The overall specification test is for the unconstrained model.

Table 6 – PMD, MLE, GMM and Optimal Instruments GMM: A Comparison

Estimates of Inflation Euler Equation: 1966:Q1 to 2001:Q4

$$z_t = (1-\mu)z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t$$

Method	Specification	μ (S.E.)	γ (S.E.)
GMM	**HP**	0.66 (0.13)	-0.055 (0.072)
GMM	**ST**	0.63 (0.13)	-0.030 (0.050)
GMM	**RULC**	0.60 (0.086)	0.053 (0.038)
MLE	**HP**	0.17 (0.037)	0.10 (0.042)
MLE	**ST**	0.18 (0.036)	0.074 (0.034)
MLE	**RULC**	0.47 (0.024)	0.050 (0.0081)
OI-GMM	**HP**	0.23 (0.093)	0.12 (0.042)
OI-GMM	**ST**	0.21 (0.11)	0.097 (0.039)
OI-GMM	**RULC**	0.45 (0.028)	0.054 (0.0081)
PMD (*h = 14*)	**HP**	0.49 (0.12)	0.050 (0.053)
PMD (*h = 14*)	**ST**	0.49 (0.12)	0.050 (0.046)
PMD (*h = 14*)	**RULC**	0.42 (0.15)	0.055 (0.057)

Unconstrained PMD (*h = 20*)

Coefficient	HP Estimate (S.E.)	ST Estimate (S.E.)	RULC Estimate (S.E.)
z_{t-1}	0.48 (0.11)	0.48 (0.12)	0.38 (0.16)
x_{t-1}	-0.02 (0.11)	-0.05 (0.12)	-0.21 (0.12)
$E_t z_{t+1}$	0.26 (0.19)	0.26 (0.18)	0.29 (0.20)
$E_t x_{t+1}$	0.09 (0.09)	0.10 (0.10)	0.21 (0.12)

Notes: z_t is a measure of inflation, x_t is a measure of the output gap, and hence economic theory would predict $\gamma > 0$. GMM, MLE and OI-GMM estimates correspond to estimates reported in Table 5 in Fuhrer and Olivei (2004). PMD estimates reported here are with impulse response horizon $h = 20$. HP refers to Hodrick-Prescott filtered log of real GDP, and ST refers to log of real GDP detrended by a deterministic segmented trend. RULC refers to real unit labor costs.

Table 7 – Overall Specification Tests of Inflation Euler Equation

Estimates of Inflation Euler Equation: 1966:Q1 to 2001:Q4

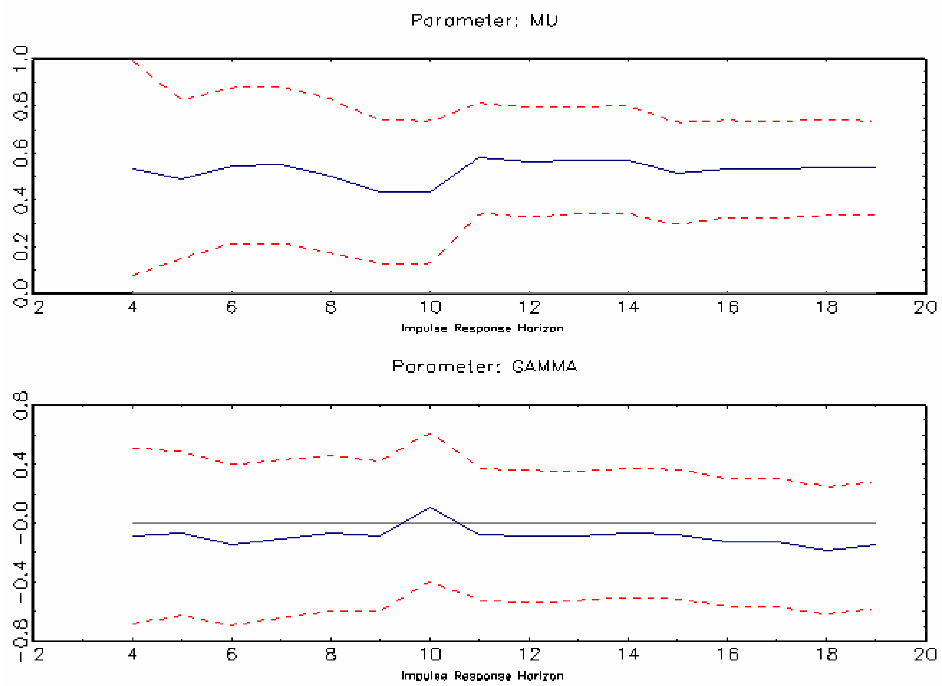
$$z_t = (1-\mu) z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t$$

	p-value of Overall Specification Test		
Horizon	**HP**	**ST**	**RULC**
4	0.772	0.771	0.987
5	0.027	0.027	0.038
6	0.031	0.043	0.005
7	0.057	0.088	0.010
8	0.065	0.099	0.017
9	0.001	0.001	0.030
10	0.001	0.001	0.038
11	0.001	0.001	0.012
12-19	0.000	0.000	0.000

Notes: p-values of the overall specification test for the unconstrained model.

Figure 1 – Parameter Estimates of the Output Euler Equation as a Function of the Impulse Response Horizon used in the First Stage Estimation

HP Detrended Ouput



Segmented Trend Output

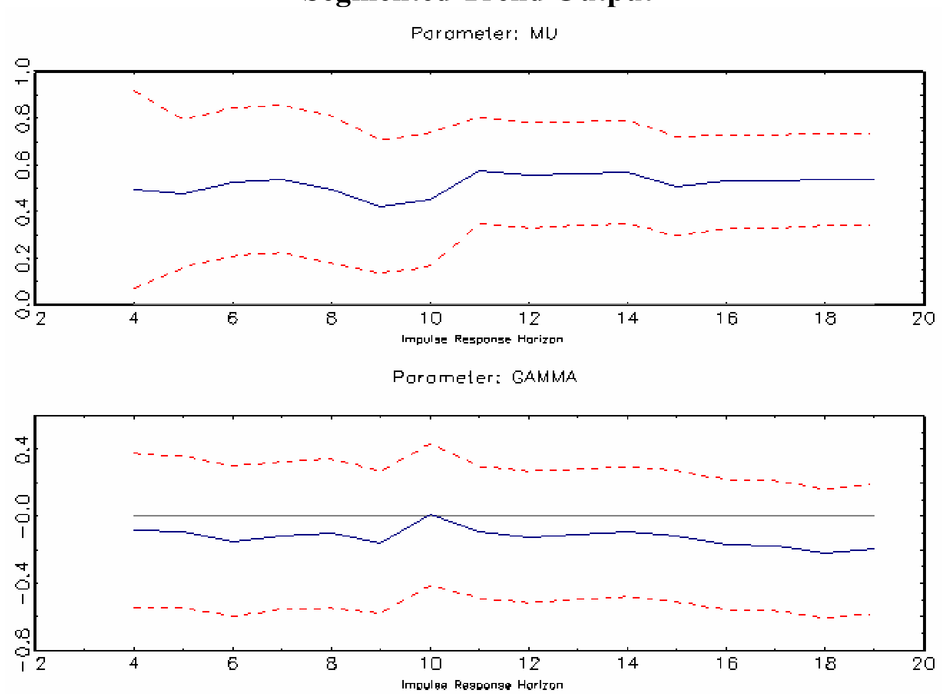
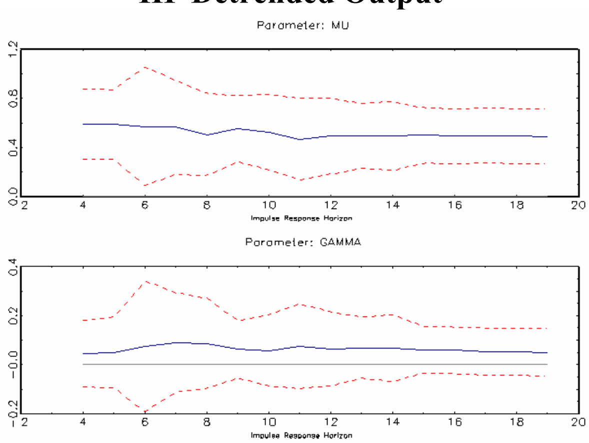
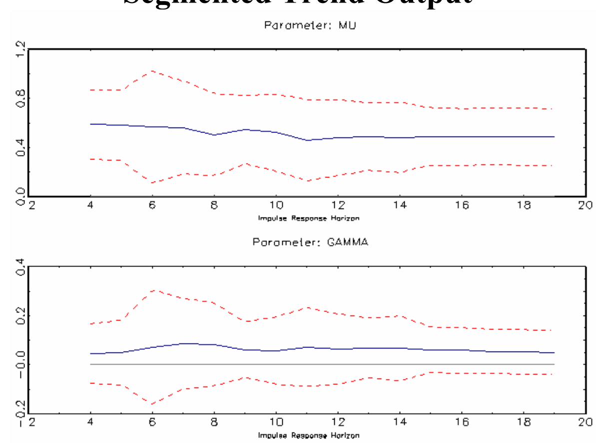


Figure 2 - Parameter Estimates of the Inflation Euler Equation as a Function of the Impulse Response Horizon used in the First Stage Estimation

HP Detrended Output



Segmented Trend Output



Real Unit Labor Costs