This version, July 2007

Estimation and Inference by the Method of Projection Minimum Distance*

Abstract

A covariance-stationary vector of variables has a Wold representation whose coefficients can be semi-parametrically estimated by local projections (Jordà, 2005). Substituting the Wold representations for variables in model expressions generates restrictions that can be used by the method of minimum distance to estimate model parameters. We call this estimator projection minimum distance (PMD) and show that its parameter estimates are consistent and asymptotically normal. In many cases, PMD is asymptotically equivalent to maximum likelihood estimation (MLE) and nests GMM as a special case. In fact, models whose ML estimation would require numerical routines (such as VARMA models) can often be estimated by simple least-squares routines and almost as efficiently by PMD. Because PMD imposes no constraints on the dynamics of the system, it is often consistent in many situations where alternative estimators would be inconsistent. We provide several Monte Carlo experiments and an empirical application in support of the new techniques introduced.

- *Keywords:* impulse response, local projection, 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 DeJong, Stephen Donald, David Drukker, Jeffrey Fuhrer, James Hamilton, Peter Hansen, Kevin Hoover, Giovanni Olivei, Paul Ruud, Frank Schorfheide, Aaron Smith, Harald Uhlig, Frank Wolak and seminar participants at the Bank of Italy, Bocconi Univeristy - IGIER, Duke University, the Federal Reserve Bank of Dallas, the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of New York, Federal Reserve Bank of San Francisco, Federal Reserve Bank of St. Louis, Southern Methodist University, Stanford University, the University of California, Berkeley, the University of California, Davis, the University of California, Riverside, the University of Houston, University of Kansas, the University of Pennsylavania, the University of Texas at Austin, and the 2006 Winter Meetings of the American Economic Association in Boston, the 2006 European Meetings of the Econometric Society in Vienna, and the 3^{rd} Macro Workshop in Vienna, 2006 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

Projection Minimum Distance (PMD) is a two-step, efficient, limited-information method of estimation based on the minimum chi-square principle (Ferguson, 1958). The first step consists in estimating the coefficients of the joint Wold representation of endogenous, exogenous and instrumental variables semi-parametrically by local projections (Jordà, 2005). We will show that these estimates are consistent and asymptotically normal even for infinite order processes and can be obtained with simple, least-squares algebra. The second step consists of minimizing the weighted quadratic distance function that relates the model's parameters with the first-step estimates. Relationships between model parameters and Wold coefficients are obtained by substituting Wold representations for variables in model expressions. The optimal weighting matrix turns out to be the inverse of the covariance matrix from the first stage. We will show that PMD estimates of the model's parameters are consistent and asymptotically normal and have good finite sample properties.

Full information techniques, such as maximum-likelihood or recent Bayesian Markov Chain-Monte Carlo approaches (see, e.g., Lubik and Schorfheide, 2004) achieve the lower efficiency information matrix bound when the model is correctly specified. Under Gaussianity, the Wold representation completely characterizes the likelihood so that, as the sample size grows to infinity, PMD approaches this lower efficiency bound as well. In fact, because covariance-stationarity implies that the Wold coefficients decay exponentially, PMD estimates are nearly fully efficient even in moderate samples. Furthermore, many models whose full-information estimates require numerical or simulation techniques for their calculation, only require two simple least-squares steps with PMD (such as estimation of vector

autoregressive, moving average models, VARMA, for example).

PMD is in the same class of limited-information estimators as GMM, M-estimators, simulated method of moments and indirect inference, to cite a few. In addition, a number of informal minimum distance estimators have been proposed to estimate dynamic macroeconomic models. We review some of these papers briefly to set our paper in context although we stress that PMD is not limited to applications in macroeconomics but rather is a general method of estimation.

Smith (1993) uses indirect inference[1] methods and simulates data from a dynamic stochastic model for different parameter values and then chooses the parameter values whose pseudo-likelihood minimizes the distance with the likelihood of a VAR estimated from the data. Naturally, the computational burden of this method is quite substantial and hence only applicable to models with relatively few parameters. Diebold, Ohanian, and Berkowitz (1998) instead minimize the distance between the spectrum implied by the model and that from the data but provide no asymptotic results and resort to the bootstrap to provide inference. Along the same lines, Kozicki and Tinsley (1999) and Sbordone (2002) extend work by Campbell and Shiller (1987, 1988) using VAR forecasts to proxy for expectations. Kozicki and Tinsley (1999) provide asymptotic properties of their recommended two-step estimator but implementation is computationally challenging in many situations. Sbordone (2002) estimates the parameters of the model by minimizing a distance function based on forecasts from a VAR. Her approach can be applied to higher dimensional problems, alas, no results are provided on the formal statistical properties of this estimator.

[1] See Gourieroux and Monfort's (1997) monograph for a more detailed discussion of indirect inference and other related simulation based estimators.

The Wold representation is sometimes referred to as the impulse response representation and the principle of minimizing the distance between the data's and the model's impulse responses has appeared in a number of recent papers, most recently in Schorfheide (2000) and Christiano et al. (2005), for example (for an earlier application see Rotemberg and Woodford, 1997). Briefly, the approach followed in this literature consists in simulating impulse responses from the economic model and then minimizing the distance between these and the impulse responses from an identified structural vector autoregression (VAR). These techniques are unsatisfactory for two main reasons. First, the success of this estimation strategy depends crucially on the ability to obtain structural impulse responses to the same fundamental shocks described by the economic model so that the minimum distance step effectively compares the same type of object. However, as Fernández-Villaverde, Rubio-Ramírez and Sargent (2005) discuss, the ability to recover the structural impulse responses of a model from a VAR is limited to very specific classes of models and depends on the ability to determine the correct method of identification of the reduced-form residual covariance matrix. Second, because it is difficult to calculate the covariance matrix of the stacked vector of impulse responses from a VAR (and to our knowledge almost never done), a suboptimal weighting matrix and simulation methods are required to estimate and report standard errors for the parameter estimates that do not have an asymptotic justification and whose statistical properties are not generally derived.

PMD is not based on identification of structural impulse responses nor does it generally consist of minimizing the distance between responses generated by the data and the model. In cross-sectional data, there is a one-to-one equivalence between PMD and GMM. In time-

series and panel data, PMD is computationally equivalent to GMM when instruments are pre-treated for serial correlation and the optimal weighting matrix is estimated by the parametric form implied by the Wold representation of the data.

Before jumping into the theoretical results, we find it useful to present our method with two simple examples in the next section. Theoretical results on consistency and asymptotic normality are provided in sections 3 and 4 for the local projection and minimum chi-square steps, respectively. Section 5 discusses the relative efficiency of PMD when compared to GMM. The performance of PMD is explored through Monte Carlo experiments in section 6 and an application in section 7. Final comments are offered in section 8.

2 Motivating Examples

This section provides the basic intuition behind PMD by stripping off the technical assumptions and derivations presented in subsequent sections. We begin with a simple time series example and then provide an example from the macroeconomics literature.

Suppose we want to estimate an ARMA(1,1) model on a sample of T observations of the variable y_t

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

Assuming y_t is covariance-stationary it has a Wold representation given by

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \tag{2}$$

where we have omitted the deterministic terms for simplicity. Substituting (2) into (1) and

matching coefficients in terms of the ε_{t-j}, we obtain the following mapping:

$$\underset{b(1,h)}{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_h \end{bmatrix}} = \underset{b(0,h-1)}{\begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_{h-1} \end{bmatrix}} \pi_1 + \underset{e}{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \theta_1$$

or more compactly $b(1,h) = b(0,h-1)\pi_1 + e\theta_1$. In section 3 we will show that an estimate of $b \equiv b(0,h)$ can be obtained from a least-squares first-stage

$$\widehat{b}_T = \left(Y'My\right)\left(y'My\right)^{-1} \tag{3}$$

where Y is a $(T-k-h)\times(h+1)$ matrix that collects the elements $Y_t = (y_t \quad y_{t+1} \quad ... \quad y_{t+h})$; y is a $(T-k-h)\times 1$ matrix collecting the elements y_t; $M = I - X(X'X)^{-1}X'$ with X a $(T-k-h)\times(k+1)$ matrix with elements $X_t = (y_{t-1} \quad ... \quad y_{t-k})$. Section 3 further shows that

$$\sqrt{(T-k-h)}\left(\widehat{b}_T - b\right) \xrightarrow{d} N\left(0, \Omega_b\right)$$

with $\widehat{\Omega}_b = \widehat{\Sigma}_v \left(y'My\right)^{-1}$,that is, the familiar least-squares result with the only wrinkle being that $\widehat{\Sigma}_v$ is an estimate of the residual variance whose specific form we will discuss in section 3.

Given the estimates $\widehat{b}_T$, an estimate of $\phi = (\pi_1 \quad \theta_1)'$ can be obtained from the second-step, minimum-distance problem

$$\min_{\phi} \widehat{Q}_T(\phi) = \mathbf{f}\left(\widehat{b}_T;\phi\right)' \widehat{W} \mathbf{f}\left(\widehat{b}_T;\phi\right) \tag{4}$$

where in this case $\mathbf{f}\left(\widehat{b}_T;\phi\right) = \left\{\widehat{b}(1,h) - \left[\widehat{b}(0,h-1)\pi_1 + e\theta_1\right]\right\}$ and $\widehat{W}$ is a weighting matrix to be described shortly. The first order conditions of this problem yield the simple least-squares result

$$\widehat{\phi}_T = -\left[\widehat{F}_\phi' \widehat{W} \widehat{F}_\phi\right]^{-1} \left[\widehat{F}_\phi' \widehat{W} \widehat{b}(1,h)\right] \tag{5}$$

where $\widehat{F}_\phi = \left(\widehat{b}\,(0,h-1) \quad e\right); \widehat{W} = \left(\widehat{F}_b \widehat{\Omega}_b \widehat{F}_b'\right)^{-1}$; and $\widehat{F}_b = (0_{h,1} \quad I_h) - \left(\widehat{\pi}_1 + \widehat{\theta}_1\right)(I_h \quad 0_{h,1})$. Given $\widehat{W}$, we show in section 4 that

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

where a convenient estimate of the covariance matrix of the parameter estimates is $\widehat{\Omega}_\phi = \left(\widehat{F}_\phi' \widehat{W} \widehat{F}_\phi\right)^{-1}$.

The two least-squares steps (3) and (5) are the essence of PMD. It is worth remarking that assuming the ε_t are $i.i.d.$ Gaussian then PMD attains the maximum likelihood efficiency lower bound when $h \to \infty$ as the sample size grows to infinity sufficiently rapidly. However, because in most covariance-stationary processes the b_j decay to zero exponentially, only a rather small value of h is necessary to quickly approach the asymptotic efficiency bound in finite samples. Secondly, PMD only requires two straight-forward least-squares steps for a model whose likelihood would require numerical techniques for its maximization. Because the method is directly scalable to vector time series, estimates for $VARMA$ models can be obtained in a computationally convenient manner. For this reason, PMD can be seen as an

alternative estimator to Hannan and Rissanen's (1982) estimator for ARMA models.

The second example that we discuss in this section is based on the hybrid New Keynesian Phillips curve presented in Galí and Gertler (1999) and which has been extensively cited in the literature (see Galí, Gertler and López-Salido, 2005 for a rebuttal of several criticisms and for additional references and citations). The basic specification in Galí and Gertler (1999) and Galí et al. (2005) is found in expression (1) of the latter paper and reproduced here with slight change of notation:

$$\pi_t = \lambda mc_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \varepsilon_{\pi t} \tag{6}$$

with $\lambda = (1-\omega)(1-\theta)(1-\beta\theta)\delta^{-1}$; $\gamma_f = \beta\theta\delta^{-1}$; $\gamma_b = \omega\delta^{-1}$; $\delta = \theta + \omega[1-\theta(1-\beta)]$; where π_t denotes inflation and mc_t log real marginal costs, both in deviations from steady-state; $1-\omega$ is the proportion of firms that optimally reset their prices, $1-\theta$ is the proportion of firms that adjust prices in a given period, and β is the discount factor. Assuming rational expectations and that $\varepsilon_{\pi t}$ is *i.i.d.*, Galí and Gertler (1999) estimate expression (6) by GMM using variables dated $t-1$ and earlier as instruments. For the purposes of the illustration, we take no position on the economic justification of the model nor on the adequacy of the estimation method given the data. Furthermore, we will only concentrate on the task of estimating the parameters λ, γ_f, and γ_b since the structural parameters ω, θ, β can then be estimated directly by classical minimum distance.[2]

Define $\mathbf{y}_{1t} = (\pi_t \quad mc_t)'$, $\mathbf{y}_{2t} = xr_t$, and hence $\mathbf{y}_t = (\mathbf{y}_{1t} \quad \mathbf{y}_{2t})'$; and $\boldsymbol{\varepsilon}_t = (\varepsilon_{\pi t} \quad \varepsilon_{mt} \quad \varepsilon_{x_t})'$. Here xr_t stands for the exchange rate, a natural predictor of inflation which appears in some

[2] That is, since $\phi = \begin{pmatrix} \lambda & \gamma_f & \gamma_b \end{pmatrix}' = g(\omega, \beta, \theta)$ and Ω_ϕ is available, then an estimate of ω, β, and θ can be obtained from the solution to the problem $\min_{\omega,\beta,\theta} (\phi - g(\omega,\beta,\theta))' \Omega_\phi^{-1} (\phi - g(\omega,\beta,\theta))$.

formulations of the Phillips curve in open economy models (see, e.g. Battini and Haldane, 1999) but not in the Galí and Gertler (1999) formulation. We use xr_t to illustrate the principle that variables omitted by the candidate model can be easily incorporated into the formulation of the PMD estimator. If $\mathbf{y}_t$ is covariance-stationary so that

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

with

$$B_j = \begin{bmatrix} b_{11}^j & b_{12}^j & b_{13}^j \\ b_{21}^j & b_{22}^j & b_{23}^j \\ b_{31}^j & b_{32}^j & b_{33}^j \end{bmatrix}$$

then it is easy to see that

$$\begin{aligned} \pi_t &= \sum_{j=0}^{\infty} b_{11}^j \varepsilon_{\pi_{t-j}} + \sum_{j=0}^{\infty} b_{12}^j \varepsilon_{m_{t-j}} + \sum_{j=0}^{\infty} b_{13}^j \varepsilon_{x_{t-j}} \\ mc_t &= \sum_{j=0}^{\infty} b_{21}^j \varepsilon_{\pi_{t-j}} + \sum_{j=0}^{\infty} b_{22}^j \varepsilon_{m_{t-j}} + \sum_{j=0}^{\infty} b_{23}^j \varepsilon_{x_{t-j}} \end{aligned} \tag{7}$$

Substituting expression (7) into the expression for the Phillips curve in (6), we obtain the following sets of conditions:

$$\begin{aligned} b_{11}^j &= \lambda b_{21}^j + \gamma_f b_{11}^{j+1} + \gamma_b b_{11}^{j-1} \quad j \geq 1 \\ b_{12}^j &= \lambda b_{22}^j + \gamma_f b_{12}^{j+1} + \gamma_b b_{12}^{j-1} \quad j > 1 \\ b_{13}^j &= \lambda b_{23}^j + \gamma_f b_{13}^{j+1} + \gamma_b b_{13}^{j-1} \quad j > 1 \end{aligned} \tag{8}$$

In order to cast the problem in terms of the minimum distance function $\mathbf{f}(b,\phi)$ of expression (4), where $b = vec(\mathbf{B})$ and $\mathbf{B} = \begin{pmatrix} I & B_1' & B_2' & \cdots & B_h' \end{pmatrix}'$, we find it useful to define $R_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$, $R_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$ and hence, the following selector matrices:

$$\begin{aligned} S_0 &= \begin{pmatrix} \mathbf{0}_{h-1,3} & (I_{h-1} \otimes R_1) & \mathbf{0}_{h-1,3} \end{pmatrix} \\ S_1 &= \begin{pmatrix} \mathbf{0}_{h-1,3} & (I_{h-1} \otimes R_2) & \mathbf{0}_{h-1,3} \end{pmatrix} \\ S_2 &= \begin{pmatrix} \mathbf{0}_{h-1,3} & \mathbf{0}_{h-1,3} & (I_{h-1} \otimes R_1) \end{pmatrix} \\ S_3 &= \begin{pmatrix} (I_{h-1} \otimes R_1) & \mathbf{0}_{h-1,3} & \mathbf{0}_{h-1,3} \end{pmatrix}. \end{aligned}$$

Thus,

$$\mathbf{f}(b,\phi) = vec\left(S_0\mathbf{B} - \left(\lambda S_1\mathbf{B} + \gamma_f S_2\mathbf{B} + \gamma_b S_3\mathbf{B}\right)\right)$$

with $\phi = \begin{pmatrix} \lambda & \gamma_f & \gamma_b \end{pmatrix}'$ and hence it reflects the distance function associated with the conditions in expression (7). Given a first stage estimator for b such that $\sqrt{T}\left(\widehat{b}_T - b_0\right) \rightarrow N(0, \Omega_b)$, then,

$$\begin{aligned} \widehat{\phi}_T &= -\left(\widehat{F}_\phi'\widehat{W}\widehat{F}_\phi\right)^{-1}\left(\widehat{F}_\phi'\widehat{W}vec\left(S_0\widehat{\mathbf{B}}\right)\right) \qquad (9) \\ \widehat{W} &= \left(\widehat{F}_b\widehat{\Omega}_b\widehat{F}_b'\right)^{-1} \\ \widehat{\Omega}_\phi &= \left(\widehat{F}_\phi'\widehat{W}\widehat{F}_\phi\right)^{-1} \end{aligned}$$

where

$$\widehat{F}_b = \left(I_3 \otimes \left(S_0 - \lambda S_1 - \gamma_f S_2 - \gamma_b S_3\right)\right)$$
$$\widehat{F}_\phi = -\left(\begin{array}{ccc} vec\left(S_1\widehat{\mathbf{B}}\right) & vec\left(S_2\widehat{\mathbf{B}}\right) & vec\left(S_3\widehat{\mathbf{B}}\right) \end{array} \right)$$

This example highlights an important feature of PMD: by recognizing that xr_t can be of predictive value for π_t, not only can one use xr_t as an instrument to estimate the parameters of interest through the impulse response coefficients of the endogenous variables $\mathbf{y}_{1t}$ with respect to the omitted variables in $\mathbf{y}_{2t}$ (as is done in the third line of expression (8)), but also the impulse response coefficients of the endogenous variables $\mathbf{y}_{1t}$ are themselves calculated so as to be orthogonal to $\mathbf{y}_{2t}$ and its lags, thus ensuring their consistency against xr (which is omitted from the formulation of the Phillips curve). We now derive the properties of our estimator.

3 First-Step: Local Projections

In this section we show that a semiparametric estimate of the Wold coefficients collected in $b = vec(\mathbf{B})$ based on local projections (Jordà, 2005) is consistent and asymptotically normal under rather general assumptions. There are several reasons why we rely on local projections rather than the more traditional inversion of a finite order VAR. First, as we will show momentarily, estimates based on local projections are consistent even for data generated by infinite order processes. This is advantageous since many macroeconomic models often have implicit reduced forms that are VARMA(p,q) representations. Second, Jordà (2005) shows that local projections are more robust (relative to VARs) to several types of misspecification. Third, the results derived here are based on *linear* local projections and hence are a

natural stepping stone for extensions based on alternative nonlinear and/or nonparametric specifications, specifications that we will investigate in a different paper and which are, for the most part, infeasible or impractical in VARs.

Local projections have the advantage of providing a simple, closed-form, analytic expression for the covariance matrix of impulse response coefficients across time and across variables. The ability to arrive at such an expression simplifies considerably the derivation of a closed-form, analytic expression for the covariance matrix of the model's parameter estimates with good efficiency properties. Expressions derived by inverting a VAR require delta method approximations and are analytically far too complex to be useful.

We begin by deriving conditions that ensure consistency of the local projection estimator and then follow with the derivation of asymptotic normality.

3.1 Consistency

Suppose

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j \boldsymbol{\varepsilon}_{t-j} \tag{10}$$

where for simplicity, but without loss of generality, we omit deterministic components (such as a constant and/or a deterministic time trend), then 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 $B(z) = \sum_{j=0}^{\infty} B_j z^j$

and the process in (10) can also be written as:

$$\mathbf{y}_t = \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t \tag{11}$$

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 on the expression that results from simple recursive substitution in this $VAR(\infty)$ representation, that is

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + \boldsymbol{\varepsilon}_{t+h} + B_1 \boldsymbol{\varepsilon}_{t+h-1} + ... + B_{h-1} \boldsymbol{\varepsilon}_{t+1} \tag{12}$$

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 (12) at lag k

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + A_k^h \mathbf{y}_{t-k+1} + \mathbf{v}_{k,t+h} \tag{13}$$

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

In what follows, we show that least squares estimates of (13) 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 consistent and asymptotically normal as long as the truncation lag grows with the sample size at an appropriate rate.

More general assumptions that allow for possible heteroskedasticity in the $\boldsymbol{\varepsilon}_t$ or on their mixing properties are possible. The key elements required are for the $\mathbf{y}_t$ to have a, possibly infinite-order, VAR representation (11) whose coefficients die-off sufficiently quickly; and for the $\boldsymbol{\varepsilon}_t$ to be sufficiently well-behaved (i.e., a white noise or a martingale difference sequence assumption) so that least-squares estimates from the truncated expression in (13) are asymptotically normal based on an appropriate law of large numbers (for a related application see, e.g. Gonçalves and Kilian, 2006). Under these more general conditions however, the ability to map the infinite VAR representation (11) into the infinite VMA representation (10) is not guaranteed. This is not a major impediment since impulse responses (understood as linear forecasts rather than conditional expectations) can still be calculated from estimates of A_1^h. On the other hand, when one assumes the $\boldsymbol{\varepsilon}_t$ are Gaussian, then PMD is asymptotically equivalent to maximum likelihood. Because we feel it is instructive to retain this point of reference (which we illustrate in the Monte Carlo exercises) and to preserve the duality between the VAR and VMA representations, we present our results in a more traditional

setting by maintaining the slightly stricter assumptions (i)-(vii) in this paper and leave more general assumptions for later research.

3.1.1 Definitions and Notation

We find it useful to define and collect the notation that we use for the derivations that follow in the remainder of the paper in this subsection. Specifically:

(i) $\underset{kr\times 1}{X_{t,k-1}} = \left(\mathbf{y}_t', \mathbf{y}_{t-1}', ..., \mathbf{y}_{t-k+1}'\right)'$

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

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

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

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

(vi) $\underset{kr\times kr}{\widehat{\Gamma}_k} = (T-k-h)^{-1} \sum_{t=k}^{T-h} X_{t,k} X_{t,k}'$

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

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

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

$$\underset{r\times kr}{\widehat{A}}(k,h) = (\widehat{A}_1^h, ..., \widehat{A}_k^h) = \widehat{\Gamma}_{1-k,h}' \widehat{\Gamma}_k^{-1} \tag{14}$$

The following theorem provides conditions under which the least-squares estimates for $A(k,h)$ are consistent.

Theorem 1 ***Consistency.*** *Let* $\{\mathbf{y}_t\}$ *satisfy (10) 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 we need, 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 (13) 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 so that,

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

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

Theorem 2 *Let* $\{\mathbf{y}_t\}$ *satisfy (10) 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* $\frac{k^3}{T}\to 0$, $k,T\to\infty$

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

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

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, 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 (10) 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,

$$\widehat{\Gamma}_{1-h|1-k}\widehat{\Gamma}(0|1-k)^{-1} = \begin{bmatrix} \widehat{A}_1^1 \\ \widehat{A}_1^2 \\ \vdots \\ \widehat{A}_1^h \end{bmatrix} = \begin{bmatrix} \widehat{B}_1 \\ \widehat{B}_2 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = \widehat{B}(1,h) \tag{15}$$

Using the usual least-squares formulas, notice that

$$\widehat{B}(1,h) = B(1,h) + \left\{ (T-k-h)^{-1} \sum_{t=k}^{T-h} \mathbf{y}_t M_{t-1,k} V_{t,h}' \right\}' \widehat{\Gamma}(0|1-k)^{-1} + o_p(1) \tag{16}$$

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 $j = 1, ..., h$. The terms vanishing in probability in (16) 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

$$\begin{aligned} &(T-k-h)^{1/2} vec\left(\widehat{B}(1,h) - B(1,h)\right) \xrightarrow{p} \\ &(T-k-h)^{1/2}\, vec\left[\left\{ (T-k-h)^{-1} \sum_{t=k}^{T-h} V_{t,h} M_{t-1,k} \mathbf{y}_t' \right\} \widehat{\Gamma}(0|1-k)^{-1} \right] \end{aligned} \tag{17}$$

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

Next notice that

$$(T-k-h)^{-1}\sum_{t=k}^{T-h} V_{t,h}V_{t,h}' \xrightarrow{p} \underset{rh\times rh}{\Sigma_v} \tag{18}$$

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

$$\underset{rh\times rh}{\Psi_B} = \begin{bmatrix} I_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ B_1 & I_r & ... & \mathbf{0}_r \\ \vdots & \vdots & ... & \vdots \\ B_{h-1} & B_{h-2} & ... & I_r \end{bmatrix}; \underset{rh\times 1}{\boldsymbol{\varepsilon}_{t,h}} = \begin{bmatrix} \boldsymbol{\varepsilon}_{t+1} \\ \vdots \\ \boldsymbol{\varepsilon}_{t+h} \end{bmatrix} \tag{19}$$

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

$$E\left[V_{t,h}V_{t,h}'\right] = \Sigma_v = \Psi_B\left(I_h \otimes \Sigma_\varepsilon\right)\Psi_B' \tag{20}$$

and therefore

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

$$\underset{r^2h\times r^2h}{\Omega_b} = \left(\underset{r\times r}{\Gamma(0|1-k)^{-1}} \otimes \underset{rh\times rh}{\Sigma_v}\right)$$

In practice, one requires sample estimates $\widehat{\Gamma}(0|1-k)^{-1}$ and $\widehat{\Sigma}_v$. With respect to the latter, notice that the parametric form of expression (20) allows us to construct a sample estimate of Ω_b by plugging-in the estimates $\widehat{B}(1,h)$ and $\widehat{\Sigma}_\varepsilon$.

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 $\mathbf{y}'_{t+j}$. 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 $\mathbf{y}_0$; the $(T-k-h)\times kr+1$ matrix $X \equiv \left(\mathbf{1}_{(T-k-h)\times 1}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k}\right)$ and the $(T-k-h)\times(T-k-h)$ matrix $M = I_{T-k-h} - X\left(X'X\right)^{-1}X'$. 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 convenient for the minimum chi-square step. Using standard properties of least-squares

$$\widehat{\mathbf{B}}_T = \widehat{B_T}(0,h) = \begin{bmatrix} I_r \\ \widehat{B}_1 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = \left[Y'M\mathbf{y}_0\right]\left[\mathbf{y}'_0M\mathbf{y}_0\right]^{-1} \tag{21}$$

with an asymptotic variance-covariance matrix for $\widehat{b}_T = vec(\widehat{\mathbf{B}}_T)$, that can be estimated with $\widehat{\Omega}_B = \left\{\left[\mathbf{y}'_0M\mathbf{y}_0\right]^{-1} \otimes \widehat{\Sigma}_v\right\}$. 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 $\widehat{\Sigma}_v = \widehat{\Psi}_B \left(I_{h+1} \otimes \widehat{\Sigma}_\epsilon \right) \widehat{\Psi}'_B$ where we extend $\widehat{\Psi}_B$ in (19) as follows

$$\underset{r(h+1)\times r(h+1)}{\widehat{\Psi}_B} = \begin{bmatrix} \mathbf{0}_r & \mathbf{0}_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ \mathbf{0}_r & I_r & \mathbf{0}_r & ... & \mathbf{0}_r \\ \mathbf{0}_r & \widehat{B}_1 & I_r & ... & \mathbf{0}_r \\ \vdots & \vdots & \vdots & ... & \vdots \\ \mathbf{0}_r & \widehat{B}_{h-1} & \widehat{B}_{h-2} & ... & I_r \end{bmatrix} \tag{22}$$

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

Thus, it is readily seen that as $h, T \to \infty$, this local projection estimator is equivalent to the maximum likelihood estimator of the Wold representation of the process $\mathbf{y}_t$ against which one could test any candidate VARMA model with a quasi-likelihood ratio test or a quasi Lagrange multiplier test. We set these issues aside since they can be recast in terms of the second step of our estimator, which we now discuss.

4 The Second Step: Minimum Chi-Square

This section begins by deriving consistency and asymptotic normality of $\widehat{\phi}_T$ obtained from the second minimum chi-square step in expression (4), 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 section 3, our objective here is to estimate ϕ by minimizing

$$\min_{\phi} \widehat{Q}_T(\phi) = \mathbf{f}\left(\widehat{b}_T;\phi\right)' \widehat{W} \mathbf{f}\left(\widehat{b}_T;\phi\right)$$

Let $Q_0(\phi)$ denote the objective function at b_0. The following theorem establishes regularity 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* $(b_0, \phi_0) = \theta_0$

(iii) The parameter space Θ *is compact*

(iv) $\mathbf{f}(b_0,\phi)$ *is continuous in a neighborhood of* $\phi_0 \in \Theta$.

(v) $\mathbf{f}\left(\widehat{b}_T;\phi\right)$ *is stochastically equicontinuous with respect to* $\widehat{b}_T$.

(vi) Instrument relevance condition: rank $[WF_\phi] = \dim(\phi)$.

(vii) Identification condition: $\dim(\mathbf{f}\left(\widehat{b}_T;\phi\right)) \geq \dim(\phi)$

Then

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

The proof is provided in the appendix. We remark that one way to derive the consistency of our estimator is to assume that h is finite (while still meeting the identification condition (vi)) even as the sample size grows to infinity. In that case, b is finite-dimensional and the proof of consistency can be done under rather standard regularity conditions. However, it is more general to assume $h, T \to \infty$ at a certain rate for h/T (an example of such a rate is given below in the proof of asymptotic normality) that will ensure that the maximum-likelihood lower efficiency bound is achieved asymptotically. Since the proof of consistency requires $\widehat{Q}_T(\phi) \xrightarrow{p} Q_0(\phi)$ uniformly, we rely on Andrews (1994, 1995), who provides results from the theory of empirical processes that allow one to verify uniform convergence when $\widehat{Q}_T(\phi)$

is stochastically equicontinuous. The conditions under which stochastic equicontinuity will hold will depend on the specific form of $\mathbf{f}(.)$ and other features of each specific application. Therefore, we prefer to state assumption (v) directly rather than stating primitive conditions that would allow one to verify stochastic equicontinuity and hence derive the proof more generically.

4.2 Asymptotic Normality

The proof of asymptotic normality relies on applying the mean value theorem to the first order conditions of the minimization of the quadratic distance function $\widehat{Q}_T(\phi)$. For this purpose, all that is required is that the weighting matrix $\widehat{W}$ converge in probability to any positive semidefinite matrix (for example, $\widehat{W} = I$). However, by choosing $\widehat{W}$ optimally, we can find the estimator with the smallest variance. This optimal choice of $\widehat{W}$ happens to be the covariance matrix of $\mathbf{f}\left(\widehat{b}_T; \phi\right)$, which results in $\widehat{Q}_T(\phi)$ having a chi-squared distribution, the essential element to derive the test of over-identifying restrictions described in the next subsection (and the basis for the minimum chi-square method of Ferguson, 1958). For these reasons, the next theorem is derived for the optimal weighting matrix instead of a generic $\widehat{W}$.

Additionally, we provide conditions that permit $h \to \infty$ with the sample size. The choice of relative rate at which $h \to \infty$ is chosen conservatively based on the literature of weak/many instruments (see Stock, Wright and Yogo, 2002, for a survey). The rate is derived such that the concentration parameter for $\widehat{b}_T$ essentially grows at the same rate as h. For this reason we need stochastic equicontinuity to hold here as well so that we can apply a central limit theorem. However, in practice, and optimal choice of h can be made with the information

criterion specifically derived for this type of problem in Hall et al. (2007).

Theorem 5 *Given the following conditions:*

(i) $\widehat{W} \xrightarrow{p} W$, *where* $W = \left(F_b \Omega_b F_b'\right)^{-1}$, *a positive semidefinite matrix with* F_b *as defined in assumption (vi) below.*

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

(v) There is a F_b *and* F_ϕ *that are continuous at* b_0 *and* ϕ_0 *respectively and*

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

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

(vi) For $F_\phi = F_\phi(\phi_0)$, *then* $F_\phi' W F_\phi$ *is invertible.*

(vii) Let $\lambda^2(h) = \widehat{b}_T'(h)(R \otimes I)\left[(R \otimes I)\Omega_b(R' \otimes I)\right]^{-1}(R' \otimes I)\widehat{b}_T(h)$, *such that* $\frac{\lambda^2(h)}{h r_1 r_2} - 1 \to \alpha > 0$ *for* $h, T \to \infty$ *and* R *a selector matrix of the appropriate columns of* $\mathbf{B}$ *given* $\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_{1t}' & \mathbf{y}_{2t}' \end{pmatrix}'$ *where* r_1 *and* r_2 *are the dimensions of* $\boldsymbol{y}_{1t}$ *and* $\mathbf{y}_{2t}$ *respectively and* $r = r_1 + r_2$.

(viii) $\mathbf{f}\left(\widehat{b}_T; \phi\right)$ *is stochastically equicontinuous with respect to* $\widehat{b}_T$.

(ix) $rank[W F_\phi] = \dim(\phi)$.

(x) $\dim\left(\mathbf{f}\left(\widehat{b}_T; \phi\right)\right) \geq \dim(\phi)$

Then:

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

where

$$\Omega_\phi = \left(F_\phi' W F_\phi\right)^{-1} \tag{23}$$

The proof is provided in the appendix. This result follows derivations similar to those for GMM and general minimum distance problems (see Newey and McFadden, 1994). The complication here is that we allow $\widehat{b}_T$ to become infinite dimensional as the sample size grows. This has two consequences: (1) to ensure asymptotic normality we have to appeal once more to empirical process theory and general stochastic equicontinuity results (see Andrews, 1994; 1995); (2) the condition $\lambda^2/(hr_1r_2) - 1 \to \alpha > 0$ is a condition on the concentration parameter of the $\widehat{b}_T$ that ensures there is sufficient explanatory power in the $\widehat{b}_T$ as $T \to \infty$ to avoid distortions in the asymptotic distribution due to weak instrument problems (see Bekker, 1994 and Staiger and Stock, 1997). In practice, one simple way to determine the optimal impulse response horizon is with the information criterion described in Hall et al. (2007). In finite samples, all asymptotic expressions (such as F_ϕ, F_b and Ω_b) can be substituted by their plug-in small sample counterparts.

We note that F_b is a function of nuisance parameters, ϕ, and therefore construction of $\widehat{W} = (F_b\Omega_b F_b')^{-1}$ in practice requires a consistently estimated $\widehat{\phi}_T$ to plug-in into the expression for $\widehat{W}$. One option is to realize that setting $\widehat{W} = I$ delivers consistent estimates of ϕ under the conditions of Theorem 4. The covariance matrix of $\widehat{\phi}_T$ with this choice of weighting matrix is not that given in expression (23) but rather:

$$\Omega_\phi = \left(F_\phi' F_\phi\right)^{-1} \left(F_\phi' F_b \Omega_b F_b' F_\phi\right) \left(F_\phi' F_\phi\right)^{-1}$$

The estimator based on the identity matrix is sometimes called the equally-weighted (EW) minimum distance estimator and sometimes it has been found to have better finite-sample properties than, for example, optimally weighted GMM estimators (see Cameron and Trivedi,

2005).

Poor small sample properties of optimally weighted minimum distance estimators are usually caused because the estimate of the optimal weighting matrix is correlated with the minimum distance function: in the case of GMM, the optimal weighting matrix is estimated as the average of the squares of the minimum distance function. In the optimally-weighted PMD estimator, consistency of the nuisance parameter $\widehat{\phi}_T$ is not required for consistency of $\widehat{b}_T$ nor $\widehat{\Omega}_b$. For this reason, any finite-sample bias will be generated by any correlation between the $\widehat{\phi}_T$ plugged into the expression for F_b, and the $\widehat{\phi}_T$ in the minimum distance function $\mathbf{f}\left(\widehat{b}_T;\widehat{\phi}_T\right)$. This, of course, is very specific to each application so a general statement is hard to make, although we have found little evidence of these issues in our Monte Carlo experiments.

Optimal-weights PMD can be obtained with a preliminary estimate of ϕ with equal-weights PMD which can then be used to construct $\widehat{F}_b$ and then redo the estimation with optimal-weights. In principle this procedure can be iterated upon, although asymptotically there is no justification to do so, and our own experiments suggest one iteration is sufficient.

4.3 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 and rank conditions require that the impulse response horizon h be chosen to guarantee that there are at least as many relevant conditions as elements in ϕ. When the number of 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.

From the proof of asymptotic normality just derived, the appendix shows that a mean-value expansion delivers the condition

$$\sqrt{T}\left(\mathbf{f}\left(\widehat{b}_T;\phi\right)-\mathbf{f}\left(b_0;\phi\right)\right)=\sqrt{T}F_b\left(\widehat{b}_T-b_0\right)+o_p\left(1\right)$$

from which

$$\sqrt{T}\left(\mathbf{f}\left(\widehat{b}_T;\phi\right)-\mathbf{f}\left(b_0;\phi\right)\right)\xrightarrow{d}N\left(\mathbf{0};F_b\Omega_bF_b'\right)$$

and hence, when $\widehat{W}$ is chosen optimally to be $\widehat{W}=\left(F_b\Omega_bF_b'\right)^{-1}$, then the minimum distance function $\widehat{Q}_T\left(\widehat{\phi}_T\right)=\mathbf{f}\left(\widehat{b}_T;\widehat{\phi}_T\right)'\widehat{W}\mathbf{f}\left(\widehat{b}_T;\widehat{\phi}_T\right)$evaluated at the optimum is a quadratic form of standardized normally distributed random variables and therefore, distributed χ^2 with degrees of freedom $\dim\left(\mathbf{f}\left(\widehat{b}_T;\phi\right)\right)-\dim\left(\phi\right)$.

4.4 PMD: A Summary for Practitioners

Consider a dynamic system characterized by an $r\times 1$ vector of variables $\mathbf{y}_t=\left(\mathbf{y}_{1t}'\quad\mathbf{y}_{2t}'\right)'$ 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 model for the variables in $\mathbf{y}_{1t}$ whose evolution can be generally summarized by a minimum distance function that relates the impulse response coefficients for $\mathbf{y}_t$ and the parameters of the model, $\mathbf{f}\left(b,\phi\right)$. Although not required, we find it helps clarify the method if we further assume that $\mathbf{f}\left(b,\phi\right)$ is of the separable form

$$\mathbf{f}(b,\phi) = \mathbf{g}(b) - \mathbf{h}(b)\,\phi$$

for $\mathbf{g}(b)$ and $\mathbf{h}(b)$ two generic functions, so that we can express all the steps in straight-forward matrix algebra.

The following steps summarize the application of PMD to this problem:

First Stage: Local Projections

1. Construct $Y = (\mathbf{y}_0, ..., \mathbf{y}_h)'$; $\mathbf{y}_0$; $X = \left(\mathbf{1}_{(T-k-h)\times r}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k}\right)$; $M = I_{(T-k-h)} - X\left(X'X\right)^{-1}X$, 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 $\widehat{b}_T = vec(\widehat{\mathbf{B}})$, where

$$\widehat{\mathbf{B}} = \left[Y'M\mathbf{y}_0\right]\left[\mathbf{y}_0'M\mathbf{y}_0\right]^{-1}$$

3. Calculate the covariance matrix of b as $\widehat{\Omega}_b = \left\{\left(\mathbf{y}_0'M\mathbf{y}_0\right)^{-1} \otimes \widehat{\Sigma}_v\right\}$, 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 (22), and $\widehat{\Sigma}_\varepsilon = \left(\widehat{\mathbf{v}}_1'\widehat{\mathbf{v}}_1\right)/(T-k-h)$; with $\widehat{\mathbf{v}}_1 = M\mathbf{y}_1 - M\mathbf{y}_0\widehat{B}_1$.

Second Stage: Minimum Chi-Square

4. The minimum distance function of the problem is

$$\widehat{Q}_T(\phi) = \mathbf{f}\left(\widehat{b}_T;\phi\right)'\widehat{W}\mathbf{f}\left(\widehat{b}_T;\phi\right)$$

The equal-weights estimator consists on setting $\widehat{W}^{EW} = I$, which can be used to obtain a preliminary estimate of ϕ

$$\widehat{\phi}_T^{EW} = \left(\mathbf{h}\left(\widehat{b}_T\right)'\mathbf{h}\left(\widehat{b}_T\right)\right)^{-1}\left(\mathbf{h}\left(\widehat{b}_T\right)'\mathbf{g}\left(\widehat{b}_T\right)\right)$$

5. Now set $\widehat{W} = \left(\widehat{F}_b \widehat{\Omega}_b \widehat{F}_b'\right)^{-1}$ where $\widehat{\Omega}_b$ has been calculated as in bullet point 3 and $\widehat{F}_b = \widehat{G}_b - \sum_{i=1}^{dim(\phi)} \widehat{H}_{i,b}\, \widehat{\phi}_{i,T}^{EW}$ where

$$\widehat{G}_b = \frac{\partial \mathbf{g}\left(\widehat{b}_T\right)}{\partial \widehat{b}_T}; \widehat{H}_{i,b} = \frac{\partial h_i\left(\widehat{b}_T\right)}{\partial \widehat{b}_T},$$

$h_i\left(\widehat{b}_T\right)$ is the i^{th} column of $\mathbf{h}\left(\widehat{b}_T\right)$ and $\widehat{\phi}_{i,T}^{EW}$ is the i^{th} element of $\widehat{\phi}_T^{EW}$. Then, the optimal-weights estimate of ϕ is

$$\widehat{\phi}_T = \left(\mathbf{h}\left(\widehat{b}_T\right)' \widehat{W} \mathbf{h}\left(\widehat{b}_T\right)\right)^{-1} \left(\mathbf{h}\left(\widehat{b}_T\right)' \widehat{W} \mathbf{g}\left(\widehat{b}_T\right)\right)$$

which can be seen as a weighted least-squares estimator, and in the more general case of a generic $\mathbf{f}(b;\phi)$, a non-linear least-squares estimator.

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

$$\widehat{\Omega}_\phi = \left(\mathbf{h}\left(\widehat{b}_T\right)' \widehat{W} \mathbf{h}\left(\widehat{b}_T\right)\right)^{-1}.$$

7. Determine the optimal impulse response horizon $\widehat{h}$ by minimizing the information criterion in Hall et al. (2007):

$$\widehat{h} = \arg \min_{h \in \{h_{\min}, \ldots, H\}} \ln\left(|\widehat{\Omega}_\phi|\right) + h \frac{\ln\left(\sqrt{T}/k\right)}{\left(\sqrt{T}/k\right)}$$

where $h_{\min}$ is such that $\dim(\mathbf{f}(\widehat{b}_T;\phi)) \geq \dim(\phi)$.

8. A test of model misspecification can be constructed as

$$\widehat{Q}_T\left(\widehat{\phi}_T\right) \xrightarrow{d} \chi^2_{\dim\left(\mathbf{f}\left(\widehat{b}_T;\phi\right)\right)-\dim(\phi)}$$

when $\dim(\mathbf{f}(\widehat{b}_T;\phi)) > \dim(\phi)$.

5 Relative Consistency and Efficiency with respect to GMM

Linear models represent a considerable portion of empirical research and conveniently simplify the expressions of GMM and PMD estimators for simpler comparison. In this context, we show that in correctly specified models with covariance-stationary data, the PMD covariance matrix appropriately corrects for serial correlation automatically whereas the GMM covariance matrix requires non-parametric solutions, such as the usual Newey-West heteroskedasticity and autocorrelation consistent covariance estimator. Such nonparametric fixes tend to have poorer properties in small samples. When the model is dynamically misspecified, PMD can provide consistent estimates of parameters of interest in situations were GMM would not be consistent. The reason is that PMD relies on instruments for endogenous variables that are orthogonal to the omitted information whereas in GMM, the instruments enter unconditionally.

Suppose we are interested in estimating the well-known canonical model

$$\mathbf{y}_t = \Phi E_t \mathbf{y}_{t+1} + \mathbf{u}_t \tag{24}$$

where $\mathbf{y}_t$ is $r \times 1$. Expression (24) can be thought of capturing a system of first-order Euler conditions from a rational expectations model, for example, although by thinking of (24) in state-space form, it is easy to see that a vast collection of models would fall under this standard set-up. More generally, we can also think of (24) as a regression with endogenous regressors. Next, we assume that the true DGP includes backward-looking terms, that is

$$\mathbf{y}_t = \Pi_1 E_t \mathbf{y}_{t+1} + \Pi_2 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{25}$$

or more generally, a regression with endogenous regressors and serial correlation. A researcher interested in estimating (24) would often use a GMM estimator with $\mathbf{y}_{t-h};\ h = 1, ...H$ as instruments. As long as $\Pi_2 = 0$, then $\widehat{\Phi}$ would a consistent estimate of Π_1, but otherwise this estimate will be biased. The reason, of course, is that the condition $E\left(\mathbf{u}_t\mathbf{y}'_{t-j}\right) = \mathbf{0}$ is violated since $E\left(\mathbf{u}_t\mathbf{y}'_{t-j}\right) = E\left(\left[\Pi_2\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t\right]\mathbf{y}'_{t-j}\right) = \Pi_2 E\left(\mathbf{y}_{t-1}\mathbf{y}'_{t-j}\right)$ which will be different from zero if $\Pi_2 \neq 0$ and $E\left(\mathbf{y}_{t-1}\mathbf{y}'_{t-j}\right) \neq 0$ – that is, $\mathbf{y}_{t-h}$ are valid instruments from the perspective of the true DGP (25) but they are invalid from the perspective of the specified model (24). One way to resolve the bias is as follows. Define $M_{t-1} = 1 - \sum_{t=2}^{T}\mathbf{y}'_{t-1}\left(\sum_{t=2}^{T}\mathbf{y}_{t-1}\mathbf{y}'_{t-1}\right)^{-1}\mathbf{y}_{t-1}$, then notice that (25) can be recast as

$$\mathbf{y}_t M_{t-1} = \Pi_1 E_t\mathbf{y}_{t+1}M_{t-1} + \boldsymbol{\varepsilon}_t M_{t-1} \tag{26}$$

since $\mathbf{y}_{t-1}M_{t-1} = 0$ by construction. Now $\mathbf{y}_{t-h};\ h = 1, ..., H$ are valid instruments for $E_t\mathbf{y}_{t+1}M_{t-1}$ since $E(\boldsymbol{\varepsilon}_t M_{t-1}\mathbf{y}'_{t-h}) = 0; h = 1, ..., H$. The lesson here is that one can estimate the parameters of a misspecified model consistently as long as the instruments are orthogonalized with respect to the possibly omitted variables. It turns out that this is exactly what differentiates GMM from PMD.

To show this, recall definitions (i)-(viii) in section 3.1.1. It is easy to see that the GMM estimator of expression (24) can be obtained from

$$\min_{\Phi} vec\left(\widehat{\Gamma}_{0-(h-1),0} - \widehat{\Gamma}_{1-h,0}\Phi\right)' \widehat{W}_T^{GMM} vec\left(\widehat{\Gamma}_{0-(h-1),0} - \widehat{\Gamma}_{1-h,0}\Phi\right) \tag{27}$$

where the optimal weighting matrix is

$$W^{GMM} = \Xi_0 + \sum_{j=1}^{\infty}(\Xi_j + \Xi_j') \tag{28}$$

with $\widehat{\Xi}_j = (T-k-h)^{-1}\left[\sum_{t=k}^{T-h} Y_{t,h}Y_{t-j,h-j}' \otimes \sum_{t=k}^{T-h} \widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_{t-j}'\right]$ and a natural way to truncate the infinite sum of the optimal weighting matrix is with a Barlett kernel as is done in Newey and West (1987), for example. It is easy to see that the GMM estimator based on (27) taking each lag $\mathbf{y}_{t-h}$ individually results in an estimate of Φ such that $E\left(\widehat{\Phi}_h^{GMM}\right) = \Pi_1 + \Pi_2\Gamma(h+1)^{-1}\Gamma(h-1)$. Thus, although $\Gamma(h) \to 0$ as $h \to \infty$ (due to the covariance-stationarity of $\mathbf{y}_t$) the term $\Gamma(h+1)^{-1}\Gamma(h-1)$ does not necessarily vanish since both the numerator and the denominator are simultaneously vanishing and the ratio is therefore indeterminate.

Consider instead the PMD estimator for this problem. First, we assume that $\mathbf{y}_t$ is covariance-stationary and has a Wold representation given by that in expression (10). Direct application of the method of undetermined coefficients on expression (24) results in the set of conditions

$$\widehat{B}\left(0, h-1\right) = \widehat{B}(1,h)\Phi$$

where from expression (15) we know that $\widehat{B}(1,h) = \widehat{\Gamma}_{1-h|1-k}\Gamma(0|1-k)^{-1}$. Consequently, the PMD estimator in this simple linear case can be directly cast as the result of minimizing

$$\min_{\Phi} vec\left(\widehat{\Gamma}_{0-(h-1)|1-k} - \widehat{\Gamma}_{1-h|1-k}\Phi\right)' \widehat{W}_T^{PMD} vec\left(\widehat{\Gamma}_{0-(h-1)|1-k} - \widehat{\Gamma}_{1-h|1-k}\Phi\right) \tag{29}$$

with

$$\begin{aligned} \widehat{W}_T^{PMD} &= \left(\widehat{\Gamma}(0|1-k)^{-1} \otimes \widehat{\Sigma}_v\right)^{-1} \\ \widehat{\Sigma}_v &= \widehat{\Psi}_B \left(I_h \otimes \widehat{\Sigma}_\varepsilon\right) \widehat{\Psi}'_B \end{aligned} \tag{30}$$

Comparing the GMM expression (27) with the PMD expression (29) it is clear that the main difference is that covariances that appear in PMD are conditional on up to k lags of the dependent variables as opposed to the unconditional covariance matrices appearing in GMM. In addition, the optimal weighting matrix in GMM is a nonparametric truncated estimate of the autocorrelation consistent covariance matrix of the instruments and the residuals whereas the PMD alternative provides a closed-form, exact expression that takes advantage of the Wold assumption. These two differences have several consequences. First, calculating the PMD bias as a function of each individual set of instruments $\mathbf{y}_{t-h}$ as we did for the GMM, results in the expression $E\left(\widehat{\Phi}_h^{PMD}\right) = \Pi_1 + \Pi_2\Gamma(h+1|k)^{-1}\Gamma(h-1|k)$. However, notice that $\Gamma(h|k) \to 0$ when either $h \to 1$ (at $h = 1$ it is exactly zero) or $h \to \infty$. Further, notice that $\Gamma(h|k) \to \Gamma(h)$ as $h \to \infty$. In fact, for $h = 1$, notice that the PMD estimator is equivalent to the GMM estimator on the model in expression (26), albeit with an efficient parametric weighting matrix. As h grows, $\widehat{\Phi}^{GMM} - \widehat{\Phi}^{PMD} \xrightarrow{p} 0$ and both estimators are equivalent. This observation suggests that a good diagnostic tool for model misspecification is to plot $\widehat{\Phi}_h^{PMD}$ (or $\widehat{\Phi}_h^{GMM} - \widehat{\Phi}_h^{PMD}$) as a function of h. This is what we do in the empirical section to the paper.

6 Monte Carlo Experiments

This section contains two types of experiments. First, we compare the finite sample properties of PMD with maximum likelihood estimation of a traditional ARMA(1,1) model. The second experiment compares PMD with GMM in the context of a traditional Euler equation. The objective is to examine the way both approaches handle biases generated by possibly omitted information and to compare the efficiency properties of both estimators in finite samples.

6.1 PMD vs. ML Estimation of ARMA Models

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 the model

$$y_t = 0.5 y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad \varepsilon_t \sim N(0, 1)$$

where θ is allowed to vary between 0 and 0.5.

Each simulation run has the following features. We use 500 burn-in observations and experiment with sample sizes $T = 50$, 100, and 400 observations. The lag-length of the

first-stage PMD estimator is determined automatically by AIC_c.[3] For the second stage, we experimented with impulse response horizons $h = 2, 5$, and 10 (although we fix $h = 5$ for the misspecification example). When $h = 2$, we have exact identification, otherwise, the model is overidentified. Impulse responses for most model specifications examined generally decay rapidly so that experiments with $h = 10$ are designed to examine the effects of including many, and possibly irrelevant conditions.

All models are estimated by MLE and 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 analytic formulas. This last computation is meant to check that the coverage implied by the analytic formulas corresponds to the Monte Carlo coverage. 1,000 Monte Carlo replications are used for each experiment.

Tables 1-4 contain the results of these experiments. Several results deserve comment. First, PMD estimates converge to the true parameter values at roughly the same pace (and sometimes faster) as MLE estimates as the sample size grows, 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 analytic standard errors that in samples bigger than 50 observations, virtually coincide with the MLE results and the Monte Carlo averages. Hence, although technically PMD achieves the MLE lower bound only asymptotically (when $h \to \infty$ as $T \to \infty$), these experiments suggest this convergence is quite rapid in finite samples. Third, we remark that MLE estimates of the ARMA(1,1) specification for some of

[3] 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.

the cases in tables 3 and 4 failed to converge due to numerical instability – the likelihood is nonlinear in the parameters and has to be optimized numerically. Rather than redoing these draws somehow, we preferred to retain the entries blank to highlight that even draws were MLE failed, PMD provided reliable estimates.

6.2 PMD vs. GMM estimation of Misspecified Models

Suppose a researcher wants to estimate the following Euler equation

$$z_t = (1-\mu)z_{t-1} + \mu E_t z_{t+1} + \gamma x_t + \varepsilon_t. \tag{31}$$

An example of such an expression in the New Keynesian hybrid Phillips curve in Galí and Gertler (1999) and is similar to the expressions we estimate in the next section based on previous work by Fuhrer and Olivei (2005). By assuming that x_t in expression (31) follows an AR(1) process, we can easily characterize the reduced-form solution as the first order VAR

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + R\varepsilon_t. \tag{32}$$

For example, when $a_{11} = a_{12} = a_{22} = 0.5$, then $\mu = 2/3$ and $\gamma = 1/3$.

Figures 1 and 2 display GMM and PMD estimates based on this model for sample sizes $T = 100$ and 400 respectively. 1,000 samples are generated with 500 burn-in observations. Each sample is then estimated by both GMM and PMD by increasing the number of instruments/horizons from two to ten. The top two panels of each figure display estimates of the parameters μ and γ respectively along with the Monte Carlo averages of the parameter

estimates for each method and the average two standard error bands. The bottom panels display the joint significance test of the h^{th} horizon impulse responses (used as a gauge of instrument significance) and the p-value of the misspecification test.

Several results deserve comment. The model is correctly specified with respect to the DGP and hence both methods provide consistent estimates of the parameters of interest. There is some slight drift in the parameter μ as the number of included instruments grows but this bias is generally rather small. We note that the joint significance test on the impulse response horizon suggests setting h to the smallest value possible (in this case 2) but even though the p-value is above 0.05 for the smaller sample, we do not observe significant distortions in the distribution. Further, higher values of h generate considerably more efficient estimates of μ and γ based on PMD relative to GMM. The p-values of the misspecification test are approximately in line with the nominal 5% value, with a slight deviation when more instruments are included. However, the size distortion is kept within 10% in any case.

6.2.1 Neglected Dynamics

To investigate the effect of neglected dynamics on the consistency properties of GMM and PMD, we experiment with a slight variation of expression (32),

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} z_{t-2} \\ x_{t-2} \end{pmatrix} + R\varepsilon_t.$$

Figures 3 and 4 examine what happens to the estimates of μ and γ now that expression (31) is misspecified with respect to this DGP whenever $b_{11} \neq 0$ (figure 3) or $b_{22} \neq 0$ (figure 4).

In figure 3 we allow b_{11} to take values in the range $[-0.5, 0.5]$, which affect the persistence of z_t and which clearly should affect estimation of the parameter μ primarily. Since the

process for x_t remains an AR(1) and exogenous with respect to the process for z_t, it is tempting to conclude that the parameter γ will be unaffected. We experiment with samples of size $T = 100$, and 300 for 1,000 replications and with models estimated with $h = 2$ both by GMM and PMD. The top panel displays biases in the estimates of μ as a function of b_{11} whereas the bottom panel displays the biases for γ instead. The most striking feature is that PMD provides virtually unbiased estimates of the coefficients μ and γ even for cases where the bias for GMM is quite substantial (such as when b_{11} approaches 0.5 and then the system has a unit root). It is interesting to note that GMM can also provide biased estimates of the parameter γ in this instance as well, although for values of b_{11} close to 0.5, PMD also presents some significant biases.

Figure 4 repeats this exercise but instead lets b_{22} vary between $[-0.5, 0.5]$. Here one would expect the reverse: very little (if any) bias in estimating μ. In fact this is what we find. Even for the extreme value of $b_{22} = -0.5$, the GMM bias is about 0.12 (PMD is essentially unbiased for any value of b_{22}). However, biases in estimating γ can be quite substantial in GMM and practically non-existent in PMD.

These results are broadly consistent with our discussion in section 2: PMD takes on an agnostic view on the underlying model that generates the data and is fully general with respect to the directions in which the proposed model is silent (in our case, the assumption that x_t is generated by an AR(1)). These Monte Carlo experiments show how omitted dynamics can easily derail traditional GMM estimates whereas PMD provides a natural and unsupervised method of adjusting previously invalid instruments for neglected serial correlation. Even when the model is correctly specified, PMD provides more efficient estimates,

the reason being that underlying the estimator is a parametric correction for serial correlation that is more effective than a traditional semiparametric Newey-West correction of the covariance matrix.

6.2.2 Omitted Information

To investigate the performance of GMM and PMD in the presence of misspecification owing to an omitted variable, we consider a variation of (31),

$$z_t = (1-\mu)z_{t-1} + \mu E_t z_{t+1} + \gamma(x_t + \delta w_t) + \epsilon_{1,t}, \tag{33}$$

where, for example, in the case of a New Keynesian Phillips curve, w_t might represent the relative price of imports. We also assume that x_t and w_t both follow AR(1) processes:

$$\begin{aligned} x_t &= a_{22}x_{t-1} + \epsilon_{2,t} \\ w_t &= a_{33}w_{t-1} + \epsilon_{3,t} \end{aligned} \tag{34}$$

Model misspecification is introduced by estimating the model under the assumption $\delta = 0$, but with data that is generated with $\delta = 0.5$. This example is meant to represent a hypothetical situation in which a NKPC for a closed economy is inappropriately applied to data for an open economy.

Table 5 summarizes the extent of bias in estimates of NKPC parameters μ and γ for PMD and GMM under this form of misspecification. In running the experiments, the shocks $\epsilon_t = [\epsilon_{1,t} \quad \epsilon_{2,t} \quad \epsilon_{3,t}]'$ were assumed to be iid $N(0, I_3)$. Results are presented for $\gamma = 0.33$, $\mu = \{0.67, 0.99\}$, $a_{22} = \{0.5, 0.8\}$, and $a_{33} = \{0.8, 0.95\}$, and for instrument lists consisting of lags of $\{z_t, x_t\}$ or $\{z_t, x_t, w_t\}$. Each estimation was based on $T = 300$ simulated observations, where an initial 500 burn-in observations were disregarded. Each experiment was repeated

for 1000 Monte Carlo repetitions. The average bias in estimates of μ and γ are presented for $h = 2$ and $h = 10$.

For $h = 2$, GMM has considerably larger bias for both μ and γ, the more persistent the omitted variable and the more forward-looking the NKPC. This result holds whether or not the instrument list includes the excluded variable w_t. When the instrument list includes w_t, PMD estimates of μ have a very small upward bias. By contrast, GMM estimates tend to have sizable downward bias.

For $h = 10$, differences between PMD and GMM are generally much smaller. Both PMD and GMM have sizable downward bias when the instrument list excludes w_t. When the instrument list includes w_t, performance of PMD tends to dominate that of GMM in terms of estimates of μ. Performance of PMD dominates GMM for estimates of γ, when μ is close to one.

7 Application: Fuhrer and Olivei (2005) 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 (2005) 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 (2005) analyze to provide the reader a context of comparison for our method. We did not explore Bayesian estimates on account that they are not reported in the Fuhrer and Olivei (2005) paper and felt that, in a large sample sense, they are covered by MLE.[4] The basic specification is (using the same notation as in Fuhrer and Olivei, 2005):

$$z_t = (1-\mu)\, z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t \tag{35}$$

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 (35), 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 (2005) 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

[4] However, we encourage the reader to check the comprehensive summary in Smets and Wouters (2003) for more details on applications of Bayesian techniques to estimation of rational expectations models.

in the GDP, chain-weighted price index. In addition, Fuhrer and Olivei (2005) 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.

Table 6 and figure 5 summarize the empirical estimates of the output Euler equation and correspond to the results in table 4 in Fuhrer and Olivei (2005), where as table 7 and figure 6 summarize the estimates of the inflation Euler equation and correspond to the results in Table 5 instead. For each Euler equation, we report the original GMM, MLE, and $OI-$GMM estimates and below these, we include the PMD results based on choosing h with Hall et al.'s (2007) information criterion. The top panels of figures 5 and 6 display the estimates of μ and γ in (35) as a function of h and the associated two-standard error bands. The bottom left panel displays the value of the information criterion (denoted with the acronym RIRSC used by Hall et al., 2007) and the bottom right panel, the p-value of the misspecification test.

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. Rather, we wish to investigate in which ways PMD coincides or departs from results that have been well studied in the literature. We begin by reviewing the estimates for the output Euler equation reported in table 6 and figure 5. The misspecification test is highly suggestive that the model is misspecified independent of how output is detrended. Nevertheless, PMD estimates of μ are very close to the MLE and OI-GMM estimates and with similar standard errors. On the other hand, PMD estimates for γ are slightly larger in magnitude, of the correct sign and statistically significant. However, while the estimates of μ appear to be rather stable to the

choice of h, we note that estimates of γ vary quite a bit as displayed in figure 5. Together with the low p-values of the misspecification test, these two pieces of evidence suggest it is best not too make strong claims on these results.

Estimates of the inflation Euler equation differ more significantly from the results in Fuhrer and Olivei (2005). For the HP filterered and ST adjusted output specifications, μ is estimated to be larger (more than double) and γ is of the wrong sign. The parameter estimates for μ and γ are rather stable to the impulse response horizon h, however. Estimates based on real unit labor costs for μ and γ are considerably closer to the MLE and OI-GMM estimates although the later does exhibit some variation as a function of h, so that some caution in staking hard claims is warranted.

With the exception of the inflation Euler model estimated with RULC, we find that the data reject most of the specifications commonly estimated (either outright, as indicated by the overidentifying restrictions test, or because of the variation of the parameter estimates as a function of h). The ability to check model specification by these two complementary methods is useful (specially in instances when the data do not reject the model but variation in parameters estimates for low values of h is substantial). With some notable exceptions, PMD estimates are often close to estimates obtained by other methods but with smaller standard errors so that at a minimum, we are able to ascertain that our results are not caused by extreme differences.

8 Conclusions

This paper introduces a disarmingly simple and novel, limited-information method of estimation. Several features make it appealing: (1) for many models, including some whose

likelihood would require numerical optimization routines, PMD only requires simple least-squares algebra; (2) for many models, PMD approximates maximum likelihood as the sample grows to infinity; (3) however, PMD is efficient in finite samples because it accounts for serial correlation parametrically; (4) as a consequence, PMD is generally more efficient than GMM; (5) PMD provides an unsupervised method of conditioning for unknown omitted dynamics that in many cases solves invalid instrument problems; (6) PMD provides many natural statistics with which to evaluate estimates of a model including, an overall misspecification test, tests on the significance of the instruments, and a way to assess which parameter estimates are most sensitive to misspecification.

The paper provides basic but generally applicable asymptotic results and ample Monte Carlo evidence in support of our claims. In addition, the empirical application provides a natural example of how PMD may be applied in practice. However, there are many research questions that space considerations prevented us from exploring. Throughout the paper, we have mentioned some of them, such as the need for a more detailed investigation of the power properties of the misspecification test in light of the GMM literature; and generalizations of our mixing and heteroskedasticity assumptions in the main theorems.

Other natural extensions include nonlinear generalizations of the local projection step to extend beyond the Wold assumption. Such generalizations are likely to be very approachable because local projections lend themselves well to more complex specifications. Similarly, we have excluded processes that are not covariance-stationary, mainly because they require slightly different assumptions on their infinite representation and the non-standard nature of the asymptotics are beyond the scope of this paper. In the end, we hope that the main con-

tribution of the paper is to provide applied researchers with a new method of estimation that is simpler than many others available, while at the same time more robust and informative.

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 $\|C\|_1^2 = \sup_{l\neq 0}\frac{l'C'Cl}{l'l}$, 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 $\frac{k}{T-k-h} \to 0$ 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

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

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

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

by recursive substitution. Thus

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

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 Define,

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

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,

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

The proof proceeds by showing that $W_{1T}\xrightarrow{p}0$, $W_{2T}\xrightarrow{p}0$, $W_{3T}\xrightarrow{p}0$.

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

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

$$\mathbf{f}\left(\widehat{b}_T; \phi\right) \xrightarrow{p} \mathbf{f}\left(b_0; \phi\right)$$

by the continuous mapping theorem since by assumption (iv), $\mathbf{f}\left(.\right)$ is continuous. Furthermore and given assumption (i)

$$\widehat{Q}_T\left(\phi\right) = \mathbf{f}\left(\widehat{b}_T; \phi\right)' \widehat{W} \mathbf{f}\left(\widehat{b}_T; \phi\right) \xrightarrow{p} \mathbf{f}\left(b_0; \phi\right)' \widehat{W} \mathbf{f}\left(b_0; \phi\right) \equiv Q_0\left(\phi\right)$$

which is a quadratic expression that is maximized at ϕ_0. Assumption (vi) provides a necessary condition for identification of the parameters (i.e., that there be at least as many 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 $\widehat{Q}_T(\phi)$.[5] Whether $\widehat{Q}_T(\phi)$ is stochastically equicontinuous depends on each application and, specifically, on the properties and assumptions made on the specific nature of $\mathbf{f}(.)$. For this reason, we directly assume here that stochastic continuity holds and we refer the reader to Andrews (1994, 1995) for examples and sets of specific conditions that apply even when b is infinite dimensional. ■

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), $\mathbf{f}(\widehat{b}_T;\phi)$ is continuously differentiable in a neighborhood of b_0 and ϕ_0 and hence $\widehat{\phi}_T$ solves the first order conditions of the minimum-distance problem

$$\min_{\phi} \mathbf{f}(\widehat{b}_T;\phi)'\widehat{W}\mathbf{f}(\widehat{b}_T;\phi)$$

[5] 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 $t \geq t_0$.

which are

$$F_\phi\left(\widehat{b}_T;\widehat{\phi}_T\right)'\widehat{W}\mathbf{f}(\widehat{b}_T;\widehat{\phi}_T)=0$$

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

$$\mathbf{f}(\widehat{b}_T;\widehat{\phi}_T)=\mathbf{f}(\widehat{b}_T;\phi_0)+F_\phi\left(\widehat{b}_T;\overline{\phi}\right)\left(\widehat{\phi}_T-\phi_0\right)$$

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

$$\mathbf{f}(\widehat{b}_T;\phi_0)=\mathbf{f}(b_0;\phi_0)+F_b\left(\overline{b};\phi_0\right)\left(\widehat{b}_T-b_0\right)$$

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

$$\begin{aligned}\sqrt{T}\mathbf{f}(\widehat{b}_T;\widehat{\phi}_T) &= \sqrt{T}\mathbf{f}(b_0;\phi_0)+F_\phi\left(\widehat{b}_T;\overline{\phi}\right)\sqrt{T}\left(\widehat{\phi}_T-\phi_0\right)+\\ &\quad F_b\left(\overline{b};\phi_0\right)\sqrt{T}\left(\widehat{b}_T-b_0\right)\end{aligned}$$

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

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

and hence

$$\sqrt{T}\mathbf{f}(\widehat{b}_T;\widehat{\phi}_T)=\sqrt{T}\mathbf{f}(b_0;\phi_0)+F_\phi\sqrt{T}\left(\widehat{\phi}_T-\phi_0\right)+F_b\sqrt{T}\left(\widehat{b}_T-b_0\right)+o_p(1)$$

In addition, by assumption (i) $\widehat{W}\xrightarrow{p}W$ and notice that $\mathbf{f}\left(b_0,\phi_0\right)=\mathbf{0}$, which combined with the first order conditions and the mean value expansions described above, allow us to write

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

Since we know that

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

then

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} -\left(F_\phi' W F_\phi\right)^{-1}\left(F_\phi' W F_b\right)\sqrt{T}\left(\widehat{b}_T - b_0\right)$$

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

$$\begin{aligned}
\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) &\xrightarrow{d} N\left(0, \Omega_\phi\right) \\
\Omega_\phi &= \left(F_\phi' W F_\phi\right)^{-1}\left(F_\phi' W F_b \Omega_b F_b' W F_\phi\right)\left(F_\phi' W F_\phi\right)^{-1}
\end{aligned}$$

Notice that since we are using the optimal weighting matrix, then $W = \left(F_b \Omega_b F_b'\right)^{-1}$ and hence, the previous expression simplifies considerably to

$$\begin{aligned}
\Omega_\phi &= \left(F_\phi' W F_\phi\right)^{-1} \\
W &= \left(F_b \Omega_b F_b'\right)^{-1}
\end{aligned}$$

■

References

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

Andrews, Donald W. K. (1994) "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62(1): 43-72.

Andrews, Donald W. K. (1995) "Non-parametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11(3): 560-596.

Battini, Nicoletta and A. G. Haldane (1999) "Forward-looking rules for Monetary Policy" in **Monetary Policy Rules**, John B. Taylor (ed.). Chicago: University of Chicago Press for NBER.

Bekker, Paul A. (1994) "Alternative Approximations to the Distribution of Instrumental Variable Estimators," *Econometrica*, 62: 657-681.

Cameron, A. Colin and Pravin K. Trivedi (2005) **Microeconometrics: Methods and Applications.** Cambridge: Cambridge University Press.

Campbell, John Y. and Robert J. Shiller (1987) "Cointegration and Tests of Present Value Models," *Journal of Political Economy,* 95: 1062-1088.

Campbell, John Y. and Robert J. Shiller (1988) "The Dividend Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies I,* 195-228.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005) "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1): 1-45.

Diebold, Francis X., Lee E. Ohanian and Jeremy Berkowitz (1998) "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies,* 65: 433-451.

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.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, and Thomas J. Sargent (2005) "A, B, C's (and D)'s for Understanding VARs," National Bureau of Economic Research, technical working paper 308.

Fuhrer, Jeffrey C. and Giovanni P. Olivei (2005) "Estimating Forward-Looking Euler Equations with GMM Estimators: An Optimal Instruments Approach," in **Models and Monetary Policy: Research in the Tradition of Dale Henderson, Richard Porter, and Peter Tinsley**, Board of Governors of the Federal Reserve System: Washington, DC, 87-104.

Galí, Jordi and Mark Gertler (1999) "Inflation Dynamics: A Structural Econometric Approach," *Journal of Monetary Economics,* 44(2): 195-222.

Galí, Jordi, Mark Gertler and David J. López-Salido (2005) "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics,* 52(6): 1107-1118.

Gonçalves, Silvia and Lutz Kilian (2006) "Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity," *Econometric Reviews, forthcoming.*

Gouriéroux, Christian and Alain Monfort (1997) **Simulation-Based Econometric Methods.** Oxford, U.K.: Oxford University Press.

Hall, Alastair, Atsushi Inoue, James M. Nason, and Barbara Rossi (2007) "Information Criteria for Impulse Response Function Matching Estimation of DSGE Models," Duke University, *mimeo.*

Hannan, Edward J. and Jorma Rissanen (1982) "Recursive Estimation of Mixed Autoregressive Moving Average Order," *Biometrika,* 69(1):81-94.

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.

Kozicki, Sharon and Peter A. Tinsley (1999) "Vector Rational Error Correction," *Journal of Economic Dynamics and Control*, 23(9-10): 1299-1327.

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.

Lubik, Thomas A. and Frank Schorfheide (2004) "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1): 190-217.

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.

Newey, Whitney K. and Kenneth D. West (1987) "A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55:703-708.

Rotemberg, Julio J. and Michael Woodford (1997) "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomics Annual*, 297-346.

Sbordone, Argia (2002) "Prices and Unit Labor Costs: Testing Models of Pricing Behavior," *Journal of Monetary Economics,* 49(2): 265-292.

Schorfheide, Frank (2000) "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15: 645-670.

Smets, Frank and Raf Wouters (2003) "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5): 1123-1175.

Smith, Anthony A. (1993) "Estimating Non-linear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics,* 8(S): S63-S84.

Staiger, Douglas and James H. Stock (1997) "Instrumental Variables Regression with Weak Instruments," *Econometrica,* 65(3): 557-586.

Stock, James H., Jonathan H. Wright and Motohiro Yogo (2002) "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics,* 20(4): 518-529.

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 – ARMA(1,1) Monte Carlo Experiments: Case (i)

$\pi_1 = 0.25$ $\Theta_1 = 0.5$ **T = 50**

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.23***	***0.49***	***0.25***	***0.44***	***0.31***	***0.28***
	SE	0.22	0.20	0.20	0.19	0.20	0.18
	SE (MC)	0.31	0.27	0.21	0.20	0.22	0.28
MLE	**Est.**	***0.22***	***0.52***	***0.23***	***0.52***	***0.22***	***0.53***
	SE	0.21	0.18	0.20	0.18	0.20	0.18
	SE (MC)	0.27	0.24	0.27	0.23	0.27	0.23

T = 100

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.24***	***0.50***	***0.25***	***0.47***	***0.27***	***0.45***
	SE	0.15	0.14	0.15	0.13	0.14	0.13
	SE (MC)	0.17	0.15	0.15	0.13	0.15	0.15
MLE	**Est.**	***0.25***	***0.51***	***0.24***	***0.51***	***0.24***	***0.50***
	SE	0.14	0.13	0.14	0.13	0.14	0.13
	SE (MC)	0.15	0.13	0.16	0.14	0.14	0.14

T = 400

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.25***	***0.51***	***0.25***	***0.50***	***0.25***	***0.50***
	SE	0.07	0.07	0.07	0.06	0.07	0.06
	SE (MC)	0.08	0.07	0.07	0.07	0.07	0.07
MLE	**Est.**	***0.25***	***0.50***	***0.25***	***0.50***	***0.24***	***0.51***
	SE	0.07	0.06	0.07	0.07	0.07	0.06
	SE (MC)	0.07	0.06	0.07	0.07	0.07	0.06

Notes: 1,000 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AICC, SE refers to the standard error calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard error based on the 1,000 estimates of the parameter. 500 burn-in observations disregarded when generating the data.

TABLE 2 – ARMA(1,1) Monte Carlo Experiments: Case (ii)

$\pi_1 = 0.5$ $\Theta_1 = 0.25$ **T = 50**

		h = 2		*h = 5*		*h = 10*	
		ρ	Θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.46***	***0.23***	***0.47***	***0.17***	***0.49***	***0.15***
	SE	0.19	0.20	0.18	0.19	0.18	0.18
	SE (MC)	0.23	0.23	0.21	0.22	0.20	0.28
MLE	**Est.**	***0.45***	***0.29***	***0.44***	***0.27***	***0.45***	***0.29***
	SE	0.20	0.20	0.20	0.21	0.20	0.20
	SE (MC)	0.21	0.23	0.23	0.25	0.19	0.22

T = 100

		h = 2		*h = 5*		*h = 10*	
		ρ	Θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.48***	***0.23***	***0.47***	***0.23***	***0.50***	***0.23***
	SE	0.13	0.14	0.13	0.14	0.12	0.13
	SE (MC)	0.15	0.16	0.14	0.16	0.13	0.18
MLE	**Est.**	***0.48***	***0.27***	***0.47***	***0.25***	***0.48***	***0.26***
	SE	0.14	0.14	0.14	0.15	0.13	0.14
	SE (MC)	0.14	0.15	0.13	0.15	0.13	0.14

T = 400

		h = 2		*h = 5*		*h = 10*	
		ρ	Θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.50***	***0.5***	***0.49***	***0.26***	***0.49***	***0.25***
	SE	0.07	0.07	0.06	0.07	0.06	0.07
	SE (MC)	0.07	0.08	0.07	0.08	0.06	0.07
MLE	**Est.**	***0.50***	***0.25***	***0.49***	***0.26***	***0.49***	***0.26***
	SE	0.07	0.07	0.07	0.07	0.07	0.07
	SE (MC)	0.06	0.07	0.07	0.07	0.06	0.07

Notes: 1,000 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AICC, SE refers to the standard error calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard error based on the 1,000 estimates of the parameter. 500 burn-in observations disregarded when generating the data.

TABLE 3 – ARMA(1,1) Monte Carlo Experiments: Case (iii)

$\pi_1 = 0$ $\Theta_1 = 0.5$ **T = 50**

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***-0.06***	***0.56***	***0.06***	***0.40***	***0.16***	***0.28***
	SE	0.36	0.32	0.27	0.25	0.25	0.22
	SE (MC)	0.61	0.55	0.28	0.29	0.31	0.37
MLE	**Est.**	**-**	**-**	**-**	**-**	**-**	**-**
	SE	-	-	-	-	-	-
	SE (MC)	-	-	-	-	-	-

T = 100

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***-0.03***	***0.54***	***0.04***	***0.45***	***0.09***	***0.41***
	SE	0.24	0.21	0.19	0.18	0.19	0.17
	SE (MC)	0.33	0.30	0.21	0.21	0.22	0.23
MLE	**Est.**	**-**	**-**	**-**	**-**	**-**	**-**
	SE	-	-	-	-	-	-
	SE (MC)	-	-	-	-	-	-

T = 400

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***-0.01***	***0.51***	***0.00***	***0.50***	***0.02***	***0.48***
	SE	0.11	0.10	0.10	0.09	0.10	0.09
	SE (MC)	0.11	0.10	0.10	0.09	0.09	0.09
MLE	**Est.**	***0.04***	***0.50***	***0.00***	***0.50***	***0.00***	***0.50***
	SE	0.10	0.09	0.10	0.09	0.10	0.08
	SE (MC)	0.10	0.09	0.10	0.09	0.09	0.08

Notes: 1,000 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AICC, SE refers to the standard error calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard error based on the 1,000 estimates of the parameter. 500 burn-in observations disregarded when generating the data.

TABLE 4 – ARMA(1,1) Monte Carlo Experiments: Case (iv)

$\pi_1 = 0.5$ $\Theta_1 = 0$ **T = 50**

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.47***	***0.04***	***0.43***	***0.03***	***0.54***	***-0.10***
	SE	0.28	0.30	0.24	0.26	0.21	0.23
	SE (MC)	0.40	0.40	0.24	0.26	0.24	0.30
MLE	**Est.**	**-**	**-**	**-**	**-**	**-**	**-**
	SE	-	-	-	-	-	-
	SE (MC)	-	-	-	-	-	-

T = 100

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.49***	***0.01***	***0.45***	***0.03***	***0.53***	***-0.04***
	SE	0.19	0.20	0.17	0.18	-.15	0.17
	SE (MC)	0.20	0.20	0.19	0.19	0.18	0.21
MLE	**Est.**	***0.49***	***-0.02***	***0.47***	***0.03***	***0.47***	***0.03***
	SE	0.17	0.20	0.18	0.20	0.18	0.20
	SE (MC)	0.18	0.20	0.19	0.20	0.18	0.20

T = 400

		h = 2		*h = 5*		*h = 10*	
		ρ	θ	ρ	θ	ρ	θ
PMD	**Est.**	***0.50***	***0.01***	***0.50***	***0.00***	***0.50***	***0.00***
	SE	0.09	0.10	0.08	0.09	0.08	0.09
	SE (MC)	0.09	0.10	0.10	0.10	0.10	0.11
MLE	**Est.**	***0.49***	***0.01***	***0.49***	***0.01***	***0.48***	***0.02***
	SE	0.09	0.10	0.09	0.10	0.09	0.10
	SE (MC)	0.09	0.10	0.09	0.10	0.09	0.10

Notes: 1,000 Monte Carlo replications, 1st-stage regression lag length chosen automatically by AICC, SE refers to the standard error calculated with the PMD/MLE formula. SE (MC) refers to the Monte Carlo standard error based on the 1,000 estimates of the parameter. 500 burn-in observations disregarded when generating the data.

TABLE 5 – Euler Equation Monte Carlo Experiments: Omitted Information

			h = 2				*h = 10*			
			Bias(μ)		*Bias(γ)*		*Bias(μ)*		*Bias(γ)*	
a_{33}	a_{22}	μ	*PMD*	*GMM*	*PMD*	*GMM*	*PMD*	*GMM*	*PMD*	*GMM*
Instruments { z_t, x_t }										
.95	**.8**	**.99**	-.34	-.52	-.24	-.38	-.48	-.48	-.33	-.32
.95	**.5**	**.99**	-.31	-.52	-.31	-.50	-.48	-.40	-.49	-.38
.80	**.8**	**.99**	-.13	-.20	-.07	-.12	-.27	-.27	-.12	-.16
.80	**.5**	**.99**	-.13	-.20	-.11	-.18	-.26	-.23	-.15	-.17
.95	**.8**	**.67**	.09	-.42	-.02	-.23	-.13	-.11	-.26	-.01
.95	**.5**	**.67**	.09	-.45	.05	-.46	-.14	-.11	-.29	-.04
.80	**.8**	**.67**	-.26	-.12	.26	.00	-.12	-.11	-.05	.00
.80	**.5**	**.67**	-.26	-.18	-.13	-.14	-.16	-.16	-.13	-.09
Instruments { z_t, x_t, w_t}										
.95	**.8**	**.99**	.04	-.33	-.05	-.24	-.26	-.41	.06	-.25
.95	**.5**	**.99**	.03	-.34	-.02	.33	-.24	-.41	.02	-.28
.80	**.8**	**.99**	.04	-.12	-.04	-.05	-.19	-.25	.06	-.14
.80	**.5**	**.99**	.04	-.11	-.01	-.10	-.13	-.19	.01	-.14
.95	**.8**	**.67**	.03	-.06	-.05	.02	-.07	-.10	.15	.02
.95	**.5**	**.67**	.06	-.06	.14	-.05	-.05	-.10	.07	-.02
.80	**.8**	**.67**	.07	-.04	-.21	.04	-.07	-.10	.14	.02
.80	**.5**	**.67**	.07	-.06	-.06	-.03	-.04	-.09	.05	-.03

Notes: 1,000 Monte Carlo replications. Sample size is T=300 with an additional initial 500 burn-in observations disregarded. Reported results are for $\gamma = 0.33$.

Table 6 – 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 ($h^* = 12$)	**HP**	0.47 (0.025)	-0.014 (0.008)
PMD ($h^* = 12$)	**ST**	0.47 (0.027)	-0.016 (0.009)

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 (2005). HP refers to Hodrick-Prescott filtered log of real GDP, and ST refers to log of real GDP detrended by a deterministic segmented trend. Optimal value of h for PMD determined by Hall et al.'s (2007) information criterion and displayed in parenthesis as h^*.

Table 7 – PMD, MLE, GMM and Optimal Instruments GMM: A Comparison

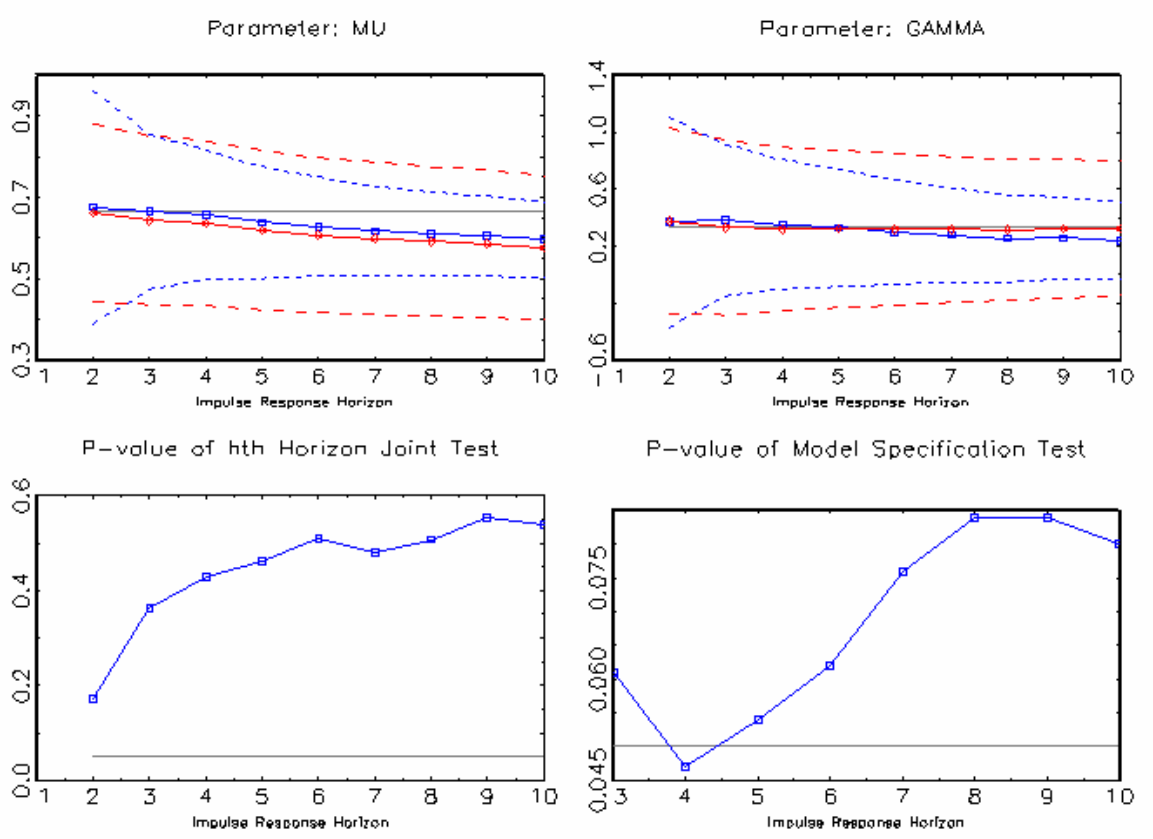
Estimates of Inflation Euler Equation: 1966:Q1 to 2001:Q4

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

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 ($h^* = 16$)	**HP**	0.59 (0.036)	-0.018 (0.019)
PMD ($h^* = 9$)	**ST**	0.63 (0.050)	-0.040 (0.019)
PMD ($h^* = 15$)	**RULC**	0.56 (0.027)	0.022 (0.010)

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 (2005). 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. Optimal value of h for PMD determined by Hall et al.'s (2007) information criterion and displayed in parenthesis as h^*.

Figure 1 - PMD and GMM Comparison when the Euler Equation is Correctly Specified. Sample Size = 100

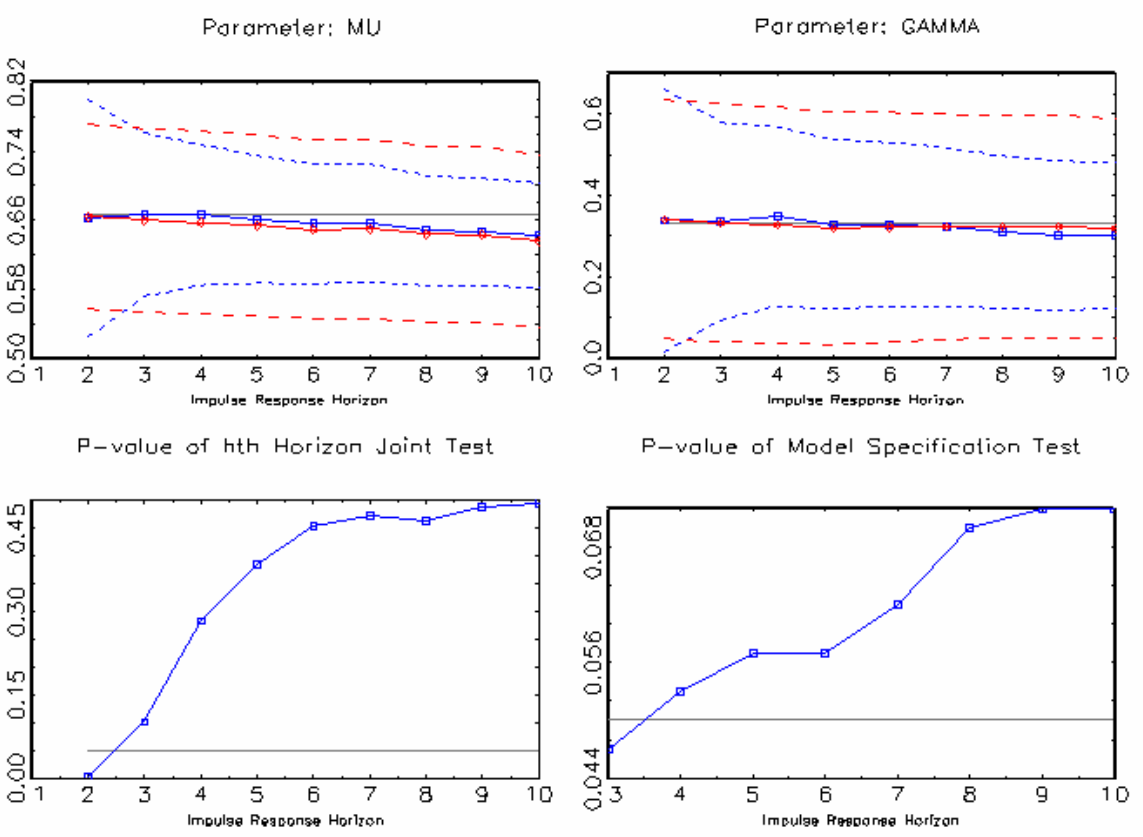


Notes: The top two panels display the Monte Carlo averages of the parameter estimates and associated two standard error bands. The bottom left panel reports the average p-value of the joint significance test on the coefficients of the h^{th} horizon whereas the bottom right panel is the power of the misspecification test at a conventional 95% level. The line with squares are the PMD estimates and the two standard error bands associated with these estimates are given by the short-dashed lines. GMM estimates are reported by the solid line with diamonds and the associated two standard error bands are given by the long-dashed lines. 1,000 Monte Carlo replications. The true parameter values are obtained by choosing :

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix};$$

which implies that $\mu = \frac{2}{3}; \gamma = \frac{1}{3}$.

Figure 2 – PMD and GMM Comparison when the Euler Equation is Correctly Specified. Sample Size = 400

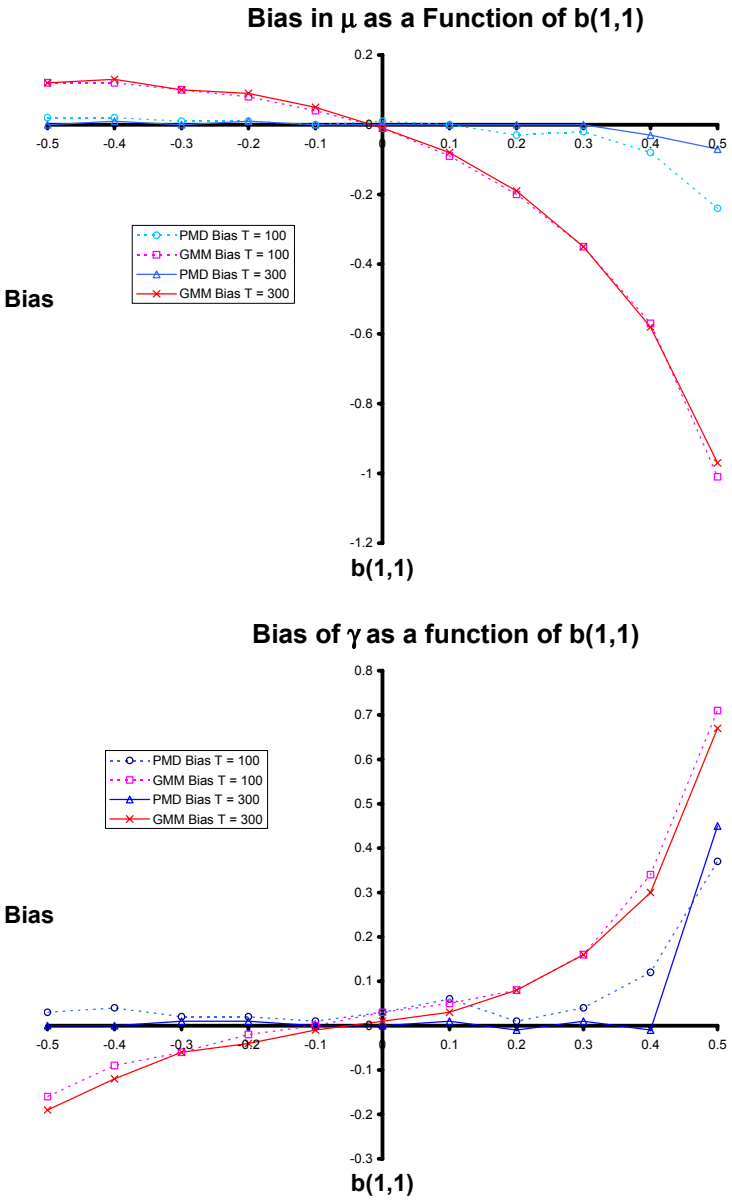


Notes: The top two panels display the Monte Carlo averages of the parameter estimates and associated two standard error bands. The bottom left panel reports the average p-value of the joint significance test on the coefficients of the h^{th} horizon whereas the bottom right panel is the power of the misspecification test at a conventional 95% level. The line with squares are the PMD estimates and the two standard error bands associated with these estimates are given by the short-dashed lines. GMM estimates are reported by the solid line with diamonds and the associated two standard error bands are given by the long-dashed lines. 1,000 Monte Carlo replications. The true parameter values are obtained by choosing :

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix};$$

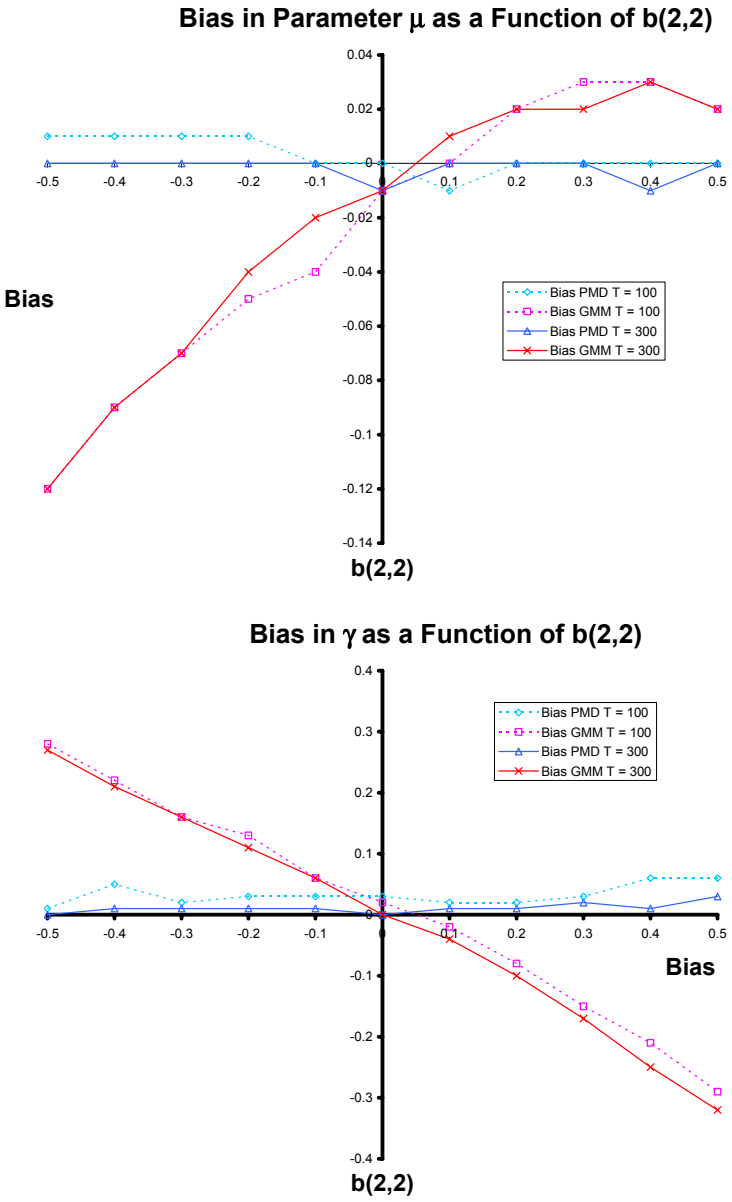
which implies that $\mu = \frac{2}{3}; \gamma = \frac{1}{3}$.

Figure 3 – PMD vs. GMM: Biases Generated by Neglected Dynamics in the Endogenous Variable



Notes: Bias generated by neglecting second order dynamics in the endogenous variable. Notice that when $b_{11} = 0.5$ or -0.5 the system has a unit root. Both PMD and GMM estimated with the first lags of the endogenous and exogenous variables only. 1,000 Monte Carlo replications.

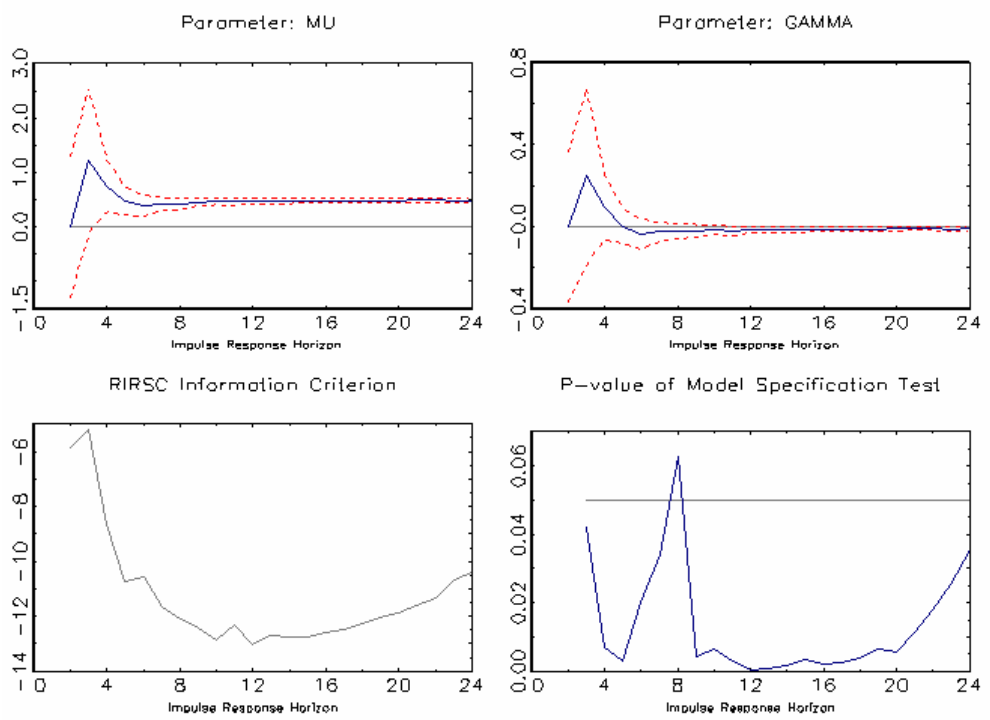
Figure 4 – PMD vs. GMM: Biases Generated by Neglected Dynamics in the Exogenous Variable



Notes: Bias generated by neglecting second order dynamics in the endogenous variable. Notice that when $b_{22} = 0.5$ or -0.5 the system has a unit root. Both PMD and GMM estimated with the first lags of the endogenous and exogenous variables only. 1,000 Monte Carlo replications.

Figure 5 – Output Euler PMD Parameter Estimates

HP Detrended Output



Segmented Trend Output

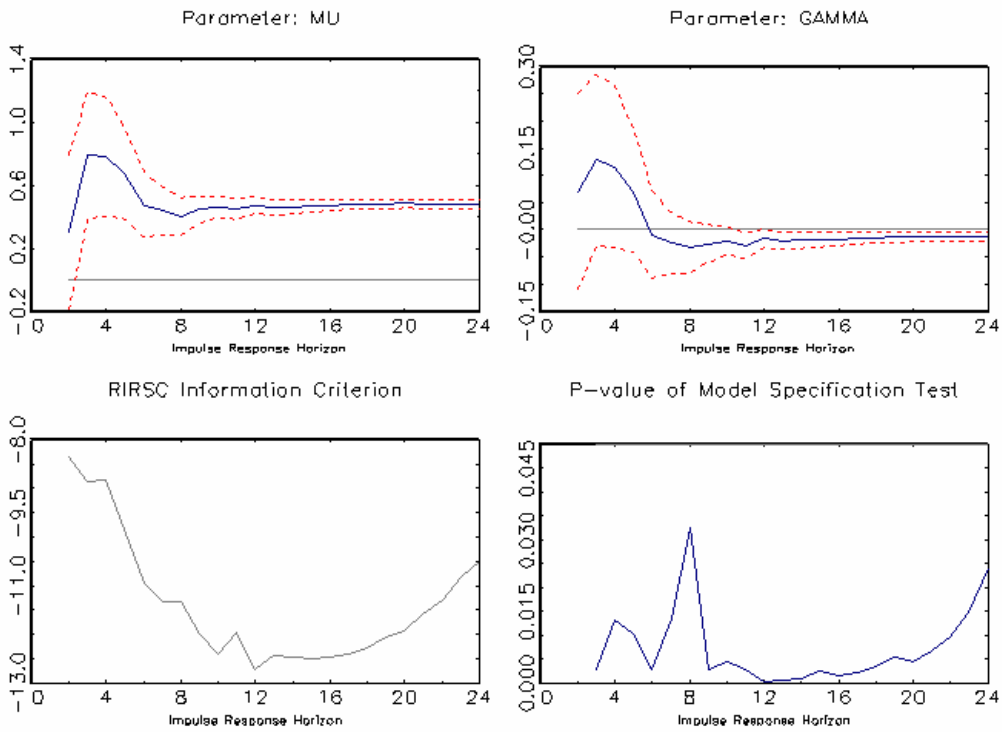
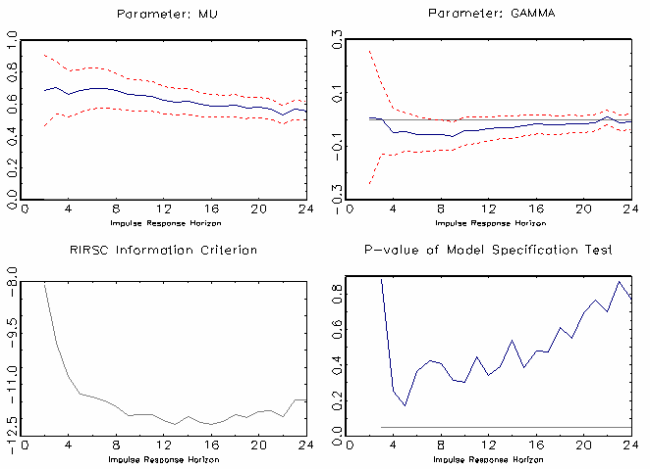
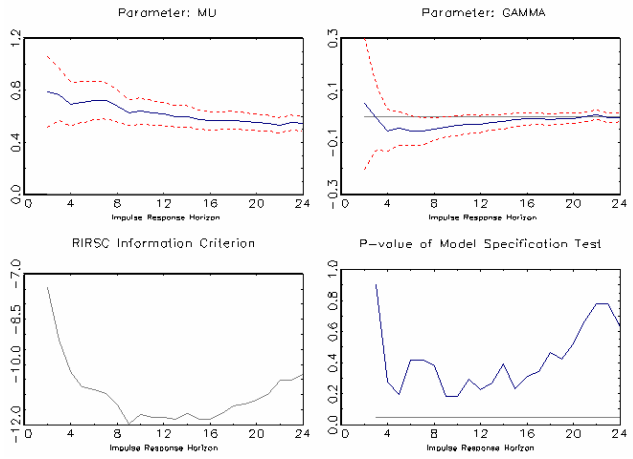


Figure 7 – Estimates of Inflation Euler Equation

HP Detrended Ouput



Segmented Trend Output



Real Unit Labor Costs