

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 semiparametrically estimated by local projections (Jordà, 2005). The parameters of a model can then be estimated from the restrictions the model imposes on the Wold coefficients by the method of minimum distance. 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 unknown model misspecification renders GMM instruments invalid. 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.

*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, David Drukker, Jeffrey Fuhrer, James Hamilton, Peter Hansen, Kevin Hoover, Giovanni Olivei, Paul Ruud, Aaron Smith, Frank Wolak and seminar participants at Duke University, the Federal Reserve Bank of New York, Federal Reserve Bank of San Francisco, Federal Reserve Bank of St. Louis, Stanford University, the University of California, Berkeley, the University of California, Davis, the University of California, Riverside, University of Kansas and the 2006 Winter Meetings of the American Economic Association in Boston, the 2006 European Meetings of the Econometric Society in Vienna, and the 3rd Macro Workshop in Vienna, 2006 for useful comments and suggestions. We are specially grateful to Harald Uhlig for comments on an earlier draft of this paper. Jordà is thankful for the hospitality of the Federal Reserve Bank of San Francisco during the preparation of this paper.

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

1 Introduction

Projection Minimum Distance (PMD) is a two-step, efficient, limited-information method of estimation based on the minimum chi-square principle (see Ferguson, 1958). In the first step, we estimate the coefficients of the Wold representation of the complete, available vector of variables semiparametrically. We will show that these estimates are consistent and asymptotically normal and can be obtained with straight-forward least-squares techniques. We then think of the equations provided by a candidate model as a multidimensional, possibly nonlinear, multiple-index function of, perhaps, only a subset of the observed and state variables in the Wold representation of the first step. This function can usually be cast to have a unique minimum where the observed and state variables can be substituted by their spectral representation (obtained in the first stage), to obtain a well-defined and unique mapping with the parameters of interest.

The second step is therefore based on minimizing the weighted quadratic distance of this function, where the optimal weights are determined by the inverse of the covariance matrix from the first step. We then prove consistency and asymptotic normality of the second step estimator. Under general conditions, this two-step estimator often provides consistent parameter estimates even when the model is misspecified with respect to the data generating process (DGP). More generally, it provides estimates that are more efficient than alternative limited information, non-likelihood based methods. The weighted quadratic objective function evaluated at the optimum has an asymptotic distribution that converges to a chi-squared and provides a test of model misspecification much along the lines of the

test for overidentifying restrictions common in minimum distance methods. We provide more details and illustrations from the time series and macroeconomic literatures below.

A simple example best demonstrates why PMD will often be preferable to other limited-information methods, such as GMM. Suppose the relation between the $T \times 1$ variable y , the $T \times r_1$ vector of endogenous variables Y , and the $T \times r_2$ vector of exogenous or predetermined variables X is generated by the expression,

$$y = Y\beta + X\gamma + u$$

for which the $T \times G$ vector Z is a vector of instruments such that $E(Zu) = 0$ and $G \geq r_2$.

Instead suppose the theory proposes the model

$$y = Y\beta + v$$

which clearly omits the X . These omitted variables could include omitted dynamics and/or other variables not directly explained by the theoretical model. The instruments Z remain natural instruments for Y , however, if $E(ZX) \neq 0$, then Z are no longer valid instruments even if $E(YX) = 0$. To see this notice that in the model specified by the researcher, $v = u + X\gamma$ and hence $E(Zv) = E(Zu) + \gamma E(ZX) \neq 0$ unless, of course, $\gamma = 0$. A natural solution is to realize that $E(Zv|X) = 0$ and hence, if $\tilde{y} = E(y|X)$; $\tilde{Y} = E(Y|X)$; and $\tilde{Z} = E(Z|X)$, then β can be estimated from the model

$$\tilde{y} = \tilde{Y}\beta + \tilde{v} \tag{1}$$

with \tilde{Z} as instruments. PMD provides an unsupervised method of instrument conditioning by semiparametrically estimating the Wold representation of the augmented vector of variables directly in the first step. Thus, the second step in PMD corresponds, in most cases (but not always), to a conditional version of the GMM step.

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 results in the exponential decay of the Wold coefficients, PMD estimates are nearly fully efficient in finite samples. In contrast, many models whose full information estimates require numerical or simulation techniques for their calculation, only require two simple least-squares steps under 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 inference¹ methods and simulates data

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

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 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, Sbordone (2002) extends work by Campbell and Shiller (1987, 1988) and 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.

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 derived.

In this regard, we would like to stress that 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. Instead, PMD minimizes a quadratic function of uniquely determined reduced-form, semiparametrically estimated impulse response coefficients and parameters that reflect underlying restrictions of the proposed model rather than its impulse responses. We now present our method and begin with a brief overview and examples.

2 Overview

2.1 Motivating Examples

In this section we provide the basic intuition behind PMD by stripping off the technical assumptions and derivations presented in subsequent sections. Suppose we want to estimate an ARMA(1,1) model on a sample of T observations of the variable y_t

$$y_t = \pi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}. \quad (2)$$

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

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \quad (3)$$

where we have omitted the deterministic terms for simplicity. Substituting (3) into (2) and matching coefficients in terms of the ε_{t-j} , we obtain the following set of conditions:

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

or more compactly

$$b(1, h) = b(0, h - 1)\pi_1 + e\theta_1. \quad (4)$$

In section 3 we will show that an estimate of $b \equiv b(0, h)$ can be obtained from a least-squares first-stage

$$\widehat{b}_T = (Y' M_z X) (X' M_z X)^{-1}$$

where Y is a $(T - k - h) \times (h + 1)$ matrix that collects the elements $Y_t = (y_{t+h} \ \dots \ y_{t+1})$; X is a $(T - k - h) \times 1$ matrix collecting the elements $X_t = y_t$; $M_z = I - Z(Z'Z)^{-1}Z'$ with

Z a $(T - k - h) \times (k + 1)$ matrix with elements $Z_t = (1 \quad y_t \quad \dots \quad y_{t-k+1})$. Section 3 further shows that

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

where

$$\widehat{\Omega}_b = \widehat{\Sigma}_v (X' M_z X)^{-1},$$

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 is also described 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

$$\min_{\phi} \widehat{Q}_T(\phi) = \mathbf{f}(\widehat{b}_T; \phi)' \widehat{W} \mathbf{f}(\widehat{b}_T; \phi) \quad (5)$$

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

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

where

$$\begin{aligned}
\widehat{F}_\phi &= \begin{pmatrix} \widehat{b}(0, h-1) & e \end{pmatrix} \\
\widehat{W} &= \left(\widehat{F}_b \widehat{\Omega}_b \widehat{F}_b' \right)^{-1} \\
\widehat{F}_b &= \begin{pmatrix} 0_{h,1} & I_h \end{pmatrix} - \left(\widehat{\pi}_1 + \widehat{\theta}_1 \right) \begin{pmatrix} I_h & 0_{h,1} \end{pmatrix}
\end{aligned}$$

Given this particular choice of weighting matrix, 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

$$\widehat{\Omega}_\phi = \left[\widehat{F}_\phi' \widehat{W} \widehat{F}_\phi \right]^{-1}.$$

These two least-squares steps are the essence of PMD. Several results deserve comment. First, it is easy to see that under the assumption that the ε_t are *i.i.d.* Gaussian then PMD attains the maximum likelihood efficiency lower bound by allowing $h \rightarrow \infty$ as the sample size grows to infinity sufficiently rapidly. However, because in most covariance-stationary processes the b_j decay exponentially to zero quite rapidly, only a rather small value of h is necessary to quickly approach the asymptotic efficiency bound. Second, we remark that our two-step PMD estimator 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. Third, in practice the optimal weighting

matrix \widehat{W} requires estimates of the parameters ϕ . These can be obtained consistently by initially setting $\widehat{W} = I_h$ (the equal-weights estimator, see Cameron and Trivedi, 2005) and then iterating the estimation with the optimal weights if desired (asymptotic arguments and Monte Carlo evidence suggests that one iteration is usually sufficient).

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:

$$\pi_t = \lambda mc_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \varepsilon_{\pi t} \quad (7)$$

with

$$\begin{aligned} \lambda &= (1 - \omega)(1 - \theta)(1 - \beta\theta)\delta^{-1} \\ \gamma_f &= \beta\theta\delta^{-1} \\ \gamma_b &= \omega\delta^{-1} \\ \delta &= \theta + \omega[1 - \theta(1 - \beta)] \end{aligned} \quad (8)$$

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 (7) 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 formulations of the Phillips curve in open economy models (see, e.g. Battini and Haldane, 1999) but not in the current 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

² That is, since expressions (8) can be collapsed into $\phi = (\lambda \quad \gamma_f \quad \gamma_b)'$ and Ω_ϕ is available, then an estimate of ω , β , and θ can be obtained from the solution to the problem $\min_{\omega, \beta, \theta} (\phi - g(\omega, \beta, \theta))' \Omega_\phi^{-1} (\phi - g(\omega, \beta, \theta))$.

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

then it is easy to see that

$$\begin{aligned} \pi_t &= \sum_{j=0}^{\infty} b_{11}^j \varepsilon_{\pi_t} + \sum_{j=0}^{\infty} b_{12}^j \varepsilon_{m_t} + \sum_{j=0}^{\infty} b_{13}^j \varepsilon_{x_t} \\ mc_t &= \sum_{j=0}^{\infty} b_{21}^j \varepsilon_{\pi_t} + \sum_{j=0}^{\infty} b_{22}^j \varepsilon_{m_t} + \sum_{j=0}^{\infty} b_{23}^j \varepsilon_{x_t} \end{aligned} \quad (9)$$

Substituting expression (9) into the expression for the Phillips curve in (7), we obtain the following sets of conditions:

$$\begin{aligned} b_{11}^j &= \lambda b_{21}^j + \gamma_f b_{11}^{j+1} + \gamma_b b_{11}^{j-1} & j \geq 1 \\ b_{12}^j &= \lambda b_{22}^j + \gamma_f b_{12}^{j+1} + \gamma_b b_{12}^{j-1} & j > 1 \\ b_{13}^j &= \lambda b_{23}^j + \gamma_f b_{13}^{j+1} + \gamma_b b_{13}^{j-1} & j > 1 \end{aligned} \quad (10)$$

In order to cast the problem in terms of the minimum distance function $\mathbf{f}(b, \phi)$ of expression (5), 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:

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

so that

$$\mathbf{f}(b, \phi) = \text{vec}(S_0 \mathbf{B} - (S_1 \mathbf{B} \ S_2 \mathbf{B} \ S_3 \mathbf{B}) \ \phi)$$

with $\phi = (\lambda \ \gamma_f \ \gamma_b)'$ reflects the distance function associated with the conditions in expression (9). Given a first stage estimator for b such that $\sqrt{T}(\hat{b}_T - b_0) \rightarrow N(0, \Omega_b)$, then,

$$\begin{aligned}
\hat{\phi}_T &= -\left(\hat{F}'_\phi \hat{W} \hat{F}_\phi\right)^{-1} \left(\hat{F}'_\phi \hat{W} \text{vec}\left(S_0 \hat{\mathbf{B}}\right)\right) \\
\hat{W} &= \left(\hat{F}_b \hat{\Omega}_b \hat{F}'_b\right)^{-1} \\
\hat{\Omega}_\phi &= \left(\hat{F}'_\phi \hat{W} \hat{F}_\phi\right)^{-1}
\end{aligned} \tag{11}$$

where

$$\begin{aligned}
\hat{F}_b &= \text{vec}\left(S_0 - \hat{\lambda} S_1 - \hat{\gamma}_f S_2 - \hat{\gamma}_b S_3\right) \\
\hat{F}_\phi &= -\left(I \otimes \begin{pmatrix} S_1 \hat{\mathbf{B}} & S_2 \hat{\mathbf{B}} & S_3 \hat{\mathbf{B}} \end{pmatrix}\right)
\end{aligned}$$

This example serves to highlight 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 (10)), 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_t (which is omitted from the formulation of the Phillips curve) along the lines of the example discussed in the introduction. This is a main avenue of departure of PMD with respect to GMM that we now discuss in more detail below.

2.2 PMD vs. GMM

This section highlights the mechanics of how PMD recovers instruments that may become invalid in a traditional GMM setting. For expositional purposes, we rely on the simplest univariate example we could think of – the lessons from this example are easily generalized to more comprehensive, multivariate settings. While we recognize there are methods to detect problematic instruments in a GMM setting, our objective is different: we want to show that PMD provides a rather general and unsupervised method of instrument management. Accordingly, suppose the DGP is characterized by the univariate backward/forward model:

$$y_t = \phi_1 E_t y_{t+1} + \phi_2 y_{t-1} + \varepsilon_t. \tag{12}$$

Instead, a candidate rational expectations model has an Euler condition given by

$$y_t = \rho E_t y_{t+1} + u_t, \quad (13)$$

which is misspecified with respect to the DGP. Based on the economic model in (13), any y_{t-j} ; $j > 1$ would be considered a valid instrument for GMM estimation and hence, an estimate of ρ would be found with the set of conditions

$$\hat{\rho}_{GMM} = \left(\frac{1}{T} \sum y_{t-j} y_{t+1} \right)^{-1} \left(\frac{1}{T} \sum y_{t-j} y_t \right). \quad (14)$$

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

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

where $\gamma_j = COV(y_t y_{t-j})$. Notice that the bias, $\phi_2 \frac{\gamma_{j-1}}{\gamma_{j+1}}$, does not disappear by selecting longer lags of y_{t-j} as instruments: although $\gamma_j \rightarrow 0$ as $j \rightarrow \infty$, $\frac{\gamma_{j-1}}{\gamma_{j+1}}$ becomes indeterminate as both the numerator and the denominator are simultaneously going to zero. Additionally, as $j \rightarrow \infty$ the correlation of the instrument with the regressor is exponentially decaying to zero – not only are these instruments invalid, they are increasingly weak. The validity of the instruments obviously depends on the specification of the DGP, not on the specification of the proposed economic model.

As long as y_t is covariance-stationary the only requirement for PMD is that the MA(∞) representation of (12) given by

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

be invertible. PMD would then first estimate the b_j by local projections

$$\hat{b}_j = \left(\frac{1}{T} \sum^T y'_{t-j} M_{t-j} y_{t-j} \right)^{-1} \left(\frac{1}{T} \sum^T y'_{t-j} M_{t-j} y_t \right) \quad (15)$$

where $M_t = 1 - z_t (z'_t z_t)^{-1} z_t$ and $z_t = (1 \ y_{t-1} \ \dots \ y_{t-k+1})$. In the second stage, notice that the mapping of the coefficients b_j and the coefficient ρ implied by the candidate model is

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

for $j \geq 1$ so that an estimate of ρ can be obtained directly from the local projections by substituting expression (15) into the previous expression and noticing that the common term $\left(\frac{1}{T} \sum^T y'_{t-j} M_{t-j} y_{t-j} \right)^{-1}$ cancels out on both sides to obtain

$$\hat{\rho}_{PMD} = \left(\frac{1}{T} \sum^T y'_{t-j} M_{t-j} y_{t+1} \right)^{-1} \left(\frac{1}{T} \sum^T y'_{t-j} M_{t-j} y_t \right) \quad (16)$$

which is the PMD counterpart to expression (14), repeated here for convenience

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

The probability limit of $\hat{\rho}_{PMD}$ is readily seen to be ϕ_1 , that is, PMD provides an unbiased estimator for the forward-looking term despite model misspecification. The explanation for this result is that PMD turns out to be equivalent to pre-treating the candidate instruments

and the endogenous variables by conditioning on the predetermined (and/or on omitted) variables, so that the pre-treated instruments become valid with respect to the conditional misspecified model. Expression (16) is equivalent to the GMM expression in (14) if one were to first regress y_t , y_{t+1} , and y_{t-j} on a constant and $y_{t-j-1}, \dots, y_{t-j-h}$ and hence replace y_t , y_{t+1} , and y_{t-j} in (14) with the residuals corresponding to the pre-treatment regressions of y_t , y_{t+1} , and y_{t-j} just described.

Not every situation will be resolved as satisfactorily with PMD. Obviously, there is no hope of addressing omitted variable bias and invalid/weak instrument problems when the omitted information is unavailable or unknown. In such situations, it is difficult to speculate whether PMD would dominate GMM. However, in general GMM relies on finding valid instruments in raw format whereas PMD can, in some cases, recover instruments that would otherwise be invalid while still providing the necessary asymptotic results. In addition, by carefully modelling the dynamic structure of the problem, we will see that PMD can be significantly more efficient than GMM. In the following sections we develop the necessary asymptotic results for PMD, beginning with the semiparametric estimator of the Wold coefficients. Then we establish the conditions under which the minimum-distance step is consistent and asymptotically normal. Monte Carlo experiments and an empirical application conclude the paper.

3 First-Step: Local Projections

In this section we show that a semiparametric estimate of the Wold coefficients collected in $b = \text{vec}(\mathbf{B})$ 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

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

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) $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Sigma_{\boldsymbol{\varepsilon}}$
 $r \times r$
- (iii) $\sum_{j=0}^{\infty} \|B_j\| < \infty$ where $\|B_j\|^2 = \text{tr}(B_j' B_j)$ and $B_0 = I_r$
- (iv) $\det \{B(z)\} \neq 0$ for $|z| \leq 1$ where $B(z) = \sum_{j=0}^{\infty} B_j z^j$

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

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

such that,

- (v) $\sum_{j=1}^{\infty} \|A_j\| < \infty$
- (vi) $A(z) = I_r - \sum_{j=1}^{\infty} A_j z^j = B(z)^{-1}$
- (vii) $\det\{A(z)\} \neq 0$ for $|z| \leq 1$.

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

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + \dots + \boldsymbol{\varepsilon}_{t+h} + B_1 \boldsymbol{\varepsilon}_{t+h-1} + \dots + B_{h-1} \boldsymbol{\varepsilon}_{t+1} \quad (19)$$

where:

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

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

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

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + \dots + A_k^h \mathbf{y}_{t-k+1} + \mathbf{v}_{k,t+h} \quad (20)$$

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

In what follows, we show that least squares estimates of (20) produce consistent estimates for A_j^h for $j = 1, \dots, k$, 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 ε_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 (18) whose coefficients die-off sufficiently quickly; and for the ε_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 (20) 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 (18) into the infinite VMA representation (17) 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 ε_t are Gaussian, we will show below that 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.

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

- (i) $X_{t,k} = (\mathbf{y}'_t, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-k+1})'$ that is, the regressors in (20).
- (ii) $\widehat{\Gamma}_{1,k,h} = (T - k - h)^{-1} \sum_{t=k}^{T-h} X_{t,k} \mathbf{y}'_{t+h}$
 $kr \times r$
- (iii) $\widehat{\Gamma}_k = (T - k - h)^{-1} \sum_{t=k}^{T-h} X_{t,k} X'_{t,k}$

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

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

The following theorem provides conditions under which the least-squares estimates for $A(k, h) = (A_1^h, \dots, A_k^h)$ are consistent.

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

- (i) $E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| < \infty$ for $1 \leq i, j, k, l \leq r$
- (ii) k is chosen as a function of T such that

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

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

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

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 (20) are asymptotically normal, although for the purposes of the PMD estimator, proving that \widehat{A}_1^h is

asymptotically normally distributed would suffice. Notice that we can write

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

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

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

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

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

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

hence

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

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

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

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

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

- (i) $E |\varepsilon_{it} \varepsilon_{jt} \varepsilon_{kt} \varepsilon_{lt}| < \infty$; $1 \leq i, j, k, l \leq r$
- (ii) k is chosen as a function of T such that $\frac{k^3}{T} \rightarrow 0$, $k, T \rightarrow \infty$
- (iii) k is chosen as a function of T such that

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

Then

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

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

$$A_T \equiv (T - k - h)^{1/2} \text{vec} \left[(T - k - h)^{-1} \left\{ \sum_{t=k}^{T-h} \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{j=1}^{h-1} B_j \boldsymbol{\varepsilon}_{t+h-j} \right) X'_{t,k} \right\} \Gamma_k^{-1} \right] \xrightarrow{d} \\ N(0, \Omega_A) \text{ with } \Omega_A = (\Gamma_k^{-1} \otimes \Sigma_h); \Sigma_h = \left(\Sigma_\varepsilon + \sum_{j=1}^{h-1} B_j \Sigma_\varepsilon B_j' \right)$$

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

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

- (i) $E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| < \infty; 1 \leq i, j, k, l \leq r$
- (ii) k is chosen as a function of T such that

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

Then

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

The proof is provided in the appendix.

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

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

- (ii) $Y_{t,h} \equiv (\mathbf{y}'_{t+1}, \dots, \mathbf{y}'_{t+h})'$
 $rh \times 1$
- (iii) $M_{t-1,k} \equiv 1 - \sum_{t=k}^{T-h} X'_{t-1,k} \left(\sum_{t=k}^{T-h} X_{t-1,k} X'_{t-1,k} \right)^{-1} X_{t-1,k}$
 1×1
- (iv) $\widehat{\Gamma}_{1|k} \equiv (T - k - h)^{-1} \sum_{t=k}^{T-h} \mathbf{y}_t M_{t-1,k} \mathbf{y}'_t$
 $r \times r$
- (v) $\widehat{\Gamma}_{1,h|k} \equiv (T - k - h)^{-1} \sum_{t=k}^{T-h} \mathbf{y}_t M_{t-1,k} Y'_{t,h}$
 $r \times rh$

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

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

Using the usual least-squares formulas, notice that

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

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

Under the conditions of theorem 2, we can write

$$(T - k - h)^{1/2} \text{vec} \left(\widehat{B}(1, h) - B(1, h) \right) \xrightarrow{p} \quad (24)$$

$$(T - k - h)^{1/2} \text{vec} \left[\left\{ (T - k - h)^{-1} \sum_{t=k}^{T-h} V_{t,h} M_{t-1,k} \mathbf{y}'_t \right\} \widehat{\Gamma}_{1|k}^{-1} \right]$$

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

Next notice that

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

The specific form of the variance-covariance matrix Σ_v can be derived as follows. Let $\mathbf{0}_j =$

$\mathbf{0}_{j \times j}$; $\mathbf{0}_{m,n} = \mathbf{0}_{m \times n}$; and recall that $V_{t,h} \equiv (\mathbf{v}'_{t+1}, \dots, \mathbf{v}'_{t+h})'$, specifically,

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

where

$$\Psi_B = \begin{bmatrix} I_r & \mathbf{0} & \dots & \mathbf{0} \\ B_1 & I_r & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ B_{h-1} & B_{h-2} & \dots & I_r \end{bmatrix}; \boldsymbol{\varepsilon}_{t,h} = \begin{bmatrix} \boldsymbol{\varepsilon}_{t+1} \\ \vdots \\ \boldsymbol{\varepsilon}_{t+h} \end{bmatrix} \quad (26)$$

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

hence

$$E [V_{t,h}V'_{t,h}] = \Sigma_v = \Psi_B (I_h \otimes \Sigma_\varepsilon) \Psi'_B \quad (27)$$

and therefore

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

$$\Omega_b = \begin{pmatrix} \Gamma_{1|k}^{-1} \otimes \Sigma_v \\ r \times r & rh \times rh \end{pmatrix}$$

In practice, one requires sample estimates $\widehat{\Gamma}_{1|k}^{-1}$ and $\widehat{\Sigma}_\varepsilon$. With respect to the latter, notice that the parametric form of expression (27) 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 $j = h, \dots, 1, 0, -1, \dots, -k$ 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 $Y \equiv (\mathbf{y}_0, \dots, \mathbf{y}_h)$; the $(T - k - h) \times r$ matrix $X \equiv \mathbf{y}_0$; the $(T - k - h) \times r(k - 1) + 1$ matrix $Z \equiv (\mathbf{1}_{(T-k-h) \times 1}, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-k+1})$ and the $(T - k - h) \times (T - k - h)$ matrix $M_z = I_{T-k-h} - Z(Z'Z)^{-1}Z'$. Notice that the inclusion of \mathbf{y}_0 in Y is a computational trick that has no other effect but to ensure that the first block of coefficients is I_r , as is required for the minimum chi-square step. Using standard properties of least-squares

$$\widehat{\mathbf{B}}_T = \widehat{B}_T(0, h) = \begin{bmatrix} I_r \\ \widehat{B}_1 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = [Y' M_z X] [X' M_z X]^{-1} \quad (28)$$

with an asymptotic variance-covariance matrix for $\widehat{b}_T = \text{vec}(\widehat{\mathbf{B}}_T)$, that can be estimated with $\widehat{\Omega}_B = \{[X' M_z X]^{-1} \otimes \widehat{\Sigma}_v\}$. 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 $\widehat{\Sigma}_v = \widehat{\Psi}_B (I_{h+1} \otimes \widehat{\Sigma}_\epsilon) \widehat{\Psi}'_B$, and by extending $\widehat{\Psi}_B$ in (26) as

$$\widehat{\Psi}_{B_{r(h+1) \times r(h+1)}} = \begin{bmatrix} \mathbf{0}_r & \mathbf{0}_r & \mathbf{0}_r & \dots & \mathbf{0}_r \\ \mathbf{0}_r & I_r & \mathbf{0}_r & \dots & \mathbf{0}_r \\ \mathbf{0}_r & \widehat{B}_1 & I_r & \dots & \mathbf{0}_r \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0}_r & \widehat{B}_{h-1} & \widehat{B}_{h-2} & \dots & I_r \end{bmatrix} \quad (29)$$

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

Thus, it is readily seen that as $h, T \rightarrow \infty$, 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 (5), 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

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

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) $\mathbf{f}(\widehat{b}_T; \phi)$ is stochastically equicontinuous.
- (vi) Instrument relevance condition: $\text{rank}[WF_{\phi}] = \dim(\phi)$.
- (vii) Identification condition: $\dim(\mathbf{f}(\widehat{b}_T; \phi)) \geq \dim(\phi)$

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 $h, T \rightarrow \infty$ 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. In such situations the proof of consistency requires $\widehat{Q}_T(\phi) \xrightarrow{p} Q_0(\phi)$ uniformly. Andrews (1994, 1995) 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 $\mathbf{f}(\cdot)$ 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 $\mathbf{f}(\widehat{b}_T; \phi)$, 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 $h \rightarrow \infty$ with the sample size. The choice of relative rate at which $h \rightarrow \infty$ 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.

Theorem 5 *Given the following conditions:*

- (i) $\widehat{W} \xrightarrow{p} W$, where $W = (F_b \Omega_b F_b')^{-1}$, 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 , $\theta = (b' \ \phi)'$
- (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 $\lambda^2(h) = \widehat{b}'_T(h) (R \otimes I) [(R \otimes I) \Omega_b (R' \otimes I)]^{-1} (R' \otimes I) \widehat{b}_T(h)$, such that $\frac{\lambda^2(h)}{hr_1r_2} - 1 \rightarrow \alpha > 0$ for $h, T \rightarrow \infty$ and R a selector matrix of the appropriate columns of \mathbf{B} given $\mathbf{y}_t = (\mathbf{y}'_{1t} \ \mathbf{y}'_{2t})'$

(viii) $\mathbf{f}(\widehat{b}_T; \phi)$ is stochastically equicontinuous.

(ix) $\text{rank}[WF_\phi] = \dim(\phi)$.

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

Then:

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

where

$$\Omega_\phi = (F'_\phi W F_\phi)^{-1} \tag{30}$$

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, 1994a, b); (2) the condition $\lambda^2/(hr_1r_2) - 1 \rightarrow \alpha > 0$ 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 $T \rightarrow \infty$ to avoid distortions in the asymptotic distribution due to weak instrument problems (see Bekker, 1994 and Staiger and Stock, 1997). In practice we advocate for a more conservative approach – a sequential joint significance test on B_j by determining h to be the $j - 1$ horizon for which the null $H_0 : B_j = 0$ is not rejected. It should be clear that

when the model has no dynamics and there are no exogenous variables \mathbf{y}_{2t} , the sequential test we propose boils down to a type of pre-test of instrument relevance. We also note that 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 (30) but rather:

$$\Omega_\phi = (F'_\phi F_\phi)^{-1} (F'_\phi F_b \Omega_b F'_b F_\phi) (F'_\phi F_\phi)^{-1}$$

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 $\mathbf{f}(\widehat{b}_T; \widehat{\phi}_T)$. This, of course, is very specific to each application so a general statement is hard to make although we have found in our Monte Carlo experiments little reason to be concerned about the optimally weighted PMD versus the equal-weights version. Hence, 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 to 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 do not suggest more than one iteration is needed.

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

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

from which

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

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

4.4 PMD: A Summary for Practitioners

Consider a dynamic system characterized by an $r \times 1$ vector of variables $\mathbf{y}_t = (\mathbf{y}'_{1t} \ \mathbf{y}'_{2t})'$ where \mathbf{y}_{1t} and \mathbf{y}_{2t} are sub-vectors of dimensions r_1 and r_2 respectively, with $r = r_1 + r_2$. A researcher specifies a model for the variables in \mathbf{y}_{1t} whose evolution can be generally summarized by a minimum distance function

$$\mathbf{f}(\mathbf{y}_{1t}, \mathbf{y}_{1t-1}, \dots, E_t \mathbf{y}_{1t+1}, \dots, E_{t+s} \mathbf{y}_{1t+s+q}; \phi) = \mathbf{f}(b, \phi)$$

where $b = \text{vec}(\mathbf{B})$, $\mathbf{B} = \left(\begin{array}{cccc} I & B_1 & \dots & B_h \end{array} \right)'$ and the \mathbf{B}_j are the Wold coefficients from the representation

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

We find it helps clarify the method if we further assume that $\mathbf{f}(b, \phi)$ is of the 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 straightforward matrix algebra although this assumption is entirely made for expositional purposes only.

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

FIRST STAGE: LOCAL PROJECTIONS

1. Construct $Y = (\mathbf{y}_0, \dots, \mathbf{y}_h)'$; $X = \mathbf{y}_0$; $Z = (\mathbf{1}_{(T-k-h) \times r}, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-k+1})$; $M_z = I_{(T-k-h)} - Z(Z'Z)^{-1}Z$, where \mathbf{y}_j is the $(T-k-h) \times r$ matrix of observations for the vector \mathbf{y}_{t+j} .
2. Compute by least squares $\hat{b}_T = \text{vec}(\hat{\mathbf{B}})$, where

$$\hat{\mathbf{B}} = [Y'M_zX][X'M_zX]^{-1}$$

3. Calculate the covariance matrix of b as $\hat{\Omega}_b = \left\{ (X'M_zX)^{-1} \otimes \hat{\Sigma}_v \right\}$, where $\hat{\Sigma}_v = \hat{\Psi}_B \left(I_h \otimes \hat{\Sigma}_\varepsilon \right) \hat{\Psi}'_B$, $\hat{\Psi}_B$ is given by expression (29), and $\hat{\Sigma}_\varepsilon = (\hat{\mathbf{v}}'_1 \hat{\mathbf{v}}_1) / (T-k-h)$; with $\hat{\mathbf{v}}_1 = M_z \mathbf{y}_1 - M_z \mathbf{y}_0 \hat{B}_1$.
4. Instrument relevance and identification conditions: Given $\hat{\Omega}_b$, construct the Wald statistic

$$W(j) = \text{vec}(\hat{B}_j)' \left(S'_j \hat{\Omega}_b S_j \right)^{-1} \text{vec}(\hat{B}_j) \xrightarrow{d} \chi_{r^2}^2$$

where S_j is a selector matrix so that $S'_j \hat{\Omega}_b S_j$ represents the covariance matrix of $\text{vec}(B_j)$. If the null hypothesis $H_0 : \text{vec}(B_j) = 0$ cannot be rejected, set $h = j - 1$,

otherwise check $j + 1$. Given this choice of h , if $\text{rank}[WF_\phi] < \dim(\phi)$ then a higher value of h is required but then estimation may suffer from a weak instrument problems that affect consistency and asymptotic normality of the parameter estimates.

SECOND STAGE: MINIMUM CHI-SQUARE

5. The minimum distance function of the problem is

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

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

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

6. 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

$$\widehat{F}_b = \widehat{G}_b - \widehat{H}_b \widehat{\phi}_T^{EW} \quad \text{where}$$

$$\frac{\partial \mathbf{g}(\widehat{b}_T)}{\partial \widehat{b}_T} = \widehat{G}_b; \quad \frac{\partial \mathbf{h}(\widehat{b}_T)}{\partial \widehat{b}_T} = \widehat{H}_b$$

Then, the optimal-weights estimate of ϕ is

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

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.

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

$$\widehat{\Omega}_\phi = \left(\widehat{F}_\phi' \widehat{W} \widehat{F}_\phi \right)^{-1}$$

where $\widehat{F}_\phi = -\mathbf{h}(\widehat{b}_T)$

8. and a test of model misspecification can be constructed as

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

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

5 Monte Carlo Experiments

This section contains two types of experiments. We begin by examining how PMD compares to maximum likelihood in the estimation of a traditional ARMA(1,1) model. We do this to highlight the efficiency properties of our estimator and to showcase the fact that the PMD estimator for this model only requires least-squares algebra even when the maximization of the likelihood requires numerical techniques. The second experiment compares the PMD estimator to GMM in the context of the estimation 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.

5.1 PMD vs. ML Estimation of ARMA Models

Although we will be experimenting with univariate $ARMA(1, 1)$ specifications, we find there is pedagogical value in discussing the more general $VARMA(1, 1)$ model first so that the reader can readily generalize the method to $VARMA(p, q)$ specifications. Accordingly, let \mathbf{y}_t be an $r \times 1$ vector that follows the following covariance-stationary process

$$\mathbf{y}_t = \underset{r \times r}{\Pi_1'} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t + \underset{r \times r}{\Theta_1'} \boldsymbol{\varepsilon}_{t-1} \quad (31)$$

with Wold decomposition,

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j' \boldsymbol{\varepsilon}_{t-j} \quad (32)$$

with $B_0 = I_r$. Substituting (32) into (31) and equating terms in $\boldsymbol{\varepsilon}_{t-j}$ the same way we did in section 2, we obtain the following conditions:

$$B_1' = I_r \Pi_1' + \Theta_1' \quad (33)$$

$$B_j' = B_{j-1}' \Pi_1' \quad \text{for } j > 1$$

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

$$S_0^* = [\mathbf{0}_{rh,r} \quad (I_h \otimes I_r)]; \quad (34)$$

$$S_1^* = [(I_h \otimes I_r) \quad \mathbf{0}_{rh,r}];$$

$$S_2^* = \begin{bmatrix} I_r & \mathbf{0}_{r,rh} \\ \mathbf{0}_{r(h-1),r(h+1)} \end{bmatrix}.$$

Defining $\mathbf{B} = B(0, h)$, then it should be clear that the conditions in (33) can be expressed as

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

so that the associated minimum-distance function is,

$$\begin{aligned} (I_r \otimes S_0^*) \widehat{b}_T - \left(I_r \otimes \begin{pmatrix} S_1^* \widehat{\mathbf{B}}_T & S_2^* \widehat{\mathbf{B}}_T \end{pmatrix} \right) \phi = \\ S^* \widehat{b}_T - g^*(\widehat{b}_T; \phi) \end{aligned} \quad (35)$$

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

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

with

$$\begin{aligned} \widehat{W}^* &= \left(\widehat{F}_b \widehat{\Omega}_b \widehat{F}_b' \right)^{-1} \\ \widehat{F}_b &= (I_r \otimes S_0^*) - \left(\widehat{\phi}' \otimes I_{rn} \right) \begin{pmatrix} I_r \otimes S_1^* \\ I_r \otimes S_2^* \end{pmatrix} \end{aligned}$$

It should be immediately obvious that once one defines the new selector matrices (34), estimation of the parameters of the model and calculation of the standard errors can be done exactly as described in section 4.

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

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

where θ is allowed to vary between 0 and 0.5. Then, to investigate the power of the misspecification test based on the overidentifying restrictions of the second stage, minimum distance step, we estimated the misspecified model

$$y_t = \rho y_{t-1} + u_t$$

by PMD.

Each simulation run has the following features. We use a burn-in of 500 observations that we then disregard to avoid initialization problems. We experiment with practical sample sizes $T = 50, 100,$ and 400 observations (for the misspecification test example, we also use $T = 200,$ and 300). 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 $h = 2, 5,$

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.

and 10 (although we fix $h = 5$ for the misspecification example). When $h = 2$, we have exact identification, otherwise, the model is overidentified. Although the impulse responses for the models we simulate generally decay within two to three periods, we experimented with $h = 10$ to examine the effects of including many additional conditions, that would seem not to include any useful information for parameter estimation.

The models in each of cases (i)-(iv) are estimated by both maximum likelihood (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 formulas. The objective is to ensure that the coverage implied by the analytical formulas corresponds to the Monte Carlo coverage. 1,000 Monte Carlo replications are used for each experiment.

Tables 1-4 contain the results for each of cases (i)-(iv). Several results deserve comment. First, PMD estimates converge to the true parameter values at roughly the same speed (sometimes faster) as MLE estimates, with estimates being close to the true values even in samples of 50 observations. However, with 50 observations, we remark some deterioration of PMD parameter estimates when $h = 10$, as would be expected by the loss of degrees of freedom. Second, PMD has 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 $h \rightarrow \infty$ as $T \rightarrow \infty$), these experiments suggest this convergence is quite rapid in practice. Third, we remark that MLE estimates of the ARMA(1,1) specification for some 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 where MLE failed, could still be estimated by PMD.

Figure 1 contains our misspecification test experiment. Because there is an extensive literature concerned with the power of the GMM test of overidentifying restrictions, we felt that similar attention should be devoted to this test and hence leave a more thorough investigation for further research. Regardless, figure 1 suggests the misspecification test has good properties. Although the misspecified model is rather close to the true model, the test has the correct size and its power increases rapidly with the sample size and as the parameter θ is allowed to grow.

Summarizing, even for relatively small samples, PMD performs comparably to MLE (with perhaps some caution when selecting large values of h in small samples). We found that the optimal weighting matrix does a good job at appropriately bringing in information from impulse responses at long horizons that may be contaminated with significant sample variation. In our experiments, parameter estimates are very stable to the choice of horizon h , the only consequence being an expected reduction in standard errors. Finally, our experiments indicate that the test of overidentifying restrictions is well behaved and can provide a suitable metric of misspecification.

5.2 PMD vs. GMM estimation of Misspecified Models

Suppose a researcher wants to estimate the following Euler equation

$$z_t = (1 - \mu)z_{t-1} + \mu E_t z_{t+1} + \gamma x_t + \varepsilon_t. \quad (36)$$

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 (36) follows an AR(1) process, we can easily characterize the reduced-form solution as the first order VAR

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + R\varepsilon_t. \quad (37)$$

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

Figures 2 and 3 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.

To investigate the effect of neglected dynamics on the consistency properties of GMM and PMD, we experiment with a slight variation of expression (37),

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

Figures 4 and 5 examine what happens to the estimates of μ and γ now that expression (36) is misspecified with respect to this DGP whenever $b_{11} \neq 0$ (figure 4) or $b_{22} \neq 0$ (figure 5).

In figure 4 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 5 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 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

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.

(using the same notation as in Fuhrer and Olivei, 2005):

$$z_t = (1 - \mu) z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t \quad (38)$$

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 (38), 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 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 5 and figure 6 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 6 and figure 7 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 as the first horizon where the joint test of significance for $h + 1$ has a p-value higher than 0.05.

The top panels of figures 6 and 7 display the estimates of μ and γ in (38) as a function of h and the associated two-standard error bands. The bottom left panel displays the p-value of the joint significance test for each horizon 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 5 and figure 6. 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 6. Together with the low p-values of the misspecification test, these two pieces of evidence suggest it is best not to make strong claims on these results.

Estimates of the inflation Euler equation differ more significantly from the results in Fuhrer and Olivei (2005). Here too we find that the misspecification test mostly rejects the model except when real unit labor costs are used instead. Interestingly, we find our estimates depart most significantly when the HP filtered and Segmented Trend adjusted output version are used – the same instances when the model is largely rejected. In these cases, γ is of

the wrong sign even if it is not statistically significant. Here too, while the estimates on μ are rather stable, estimates of γ vary quite a bit as a function of h , although they tend to be almost always negative. Estimates based on real unit labor costs, for which the model is not rejected by our test, suggest μ attains a slightly higher value than MLE or OI-GMM instruments (0.67 for the former versus 0.47 and 0.45 for the latter two) but the coefficient of γ is very similar (0.033 versus 0.050 and 0.054). Since this estimate is borderline significant and does exhibit some variation as a function of h , some caution in staking hard claims is warranted.

Several conclusions can be made from this exercise. First, these hybrid Euler specifications are mostly rejected by the data. This complicates any comparison between methods but we are heartened by the similarities we find when the misspecification test is not rejected and also in regard to the estimates of μ and its standard errors. Second, PMD provides useful information to assess the dimensions in which the model may be failing. In addition to the overall misspecification test, we find that the variability of the parameter estimates with h is helpful in uncovering which parameter estimates are likely to be problematic. Third, we feel that reporting the p-value of the joint test of significance for h provides information that lets the end user assess the quality of the instruments used in an intuitive way.

7 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 approximately as efficient in small 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 contribution 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.

8 Appendix

Proof. Theorem 1

Notice that

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

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,

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

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

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

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

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

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)

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

by independence. Hence

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

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

Next, consider $\|U_{3T}\| \xrightarrow{p} 0$. The proof is very similar since ε_{t+h-j} , $j = 1, \dots, h-1$ 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

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

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

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

Hence

$$k^{1/2} \sum_{j=k+1}^{\infty} \|A_j^h\| = k^{1/2} \left\{ \sum_{j=k+1}^{\infty} \|B_{h-1}A_j + B_{h-2}A_{j+1} + \dots + B_1A_{j+h-2} + A_{j+h-1}\| \right\}$$

by recursive substitution. Thus

$$k^{1/2} \sum_{j=k+1}^{\infty} \|A_j^h\| \leq k^{1/2} \left\{ \sum_{j=k+1}^{\infty} \|B_{h-1}A_j\| + \dots + \|B_1A_{j+h-2}\| + \|A_{j+h-1}\| \right\}$$

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

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

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

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

instead. ■

Proof. Theorem 2

We begin by showing that $W_{1T} \xrightarrow{p} 0$. Lewis and Reinsel (1985) show that under assumption (ii), $k^{1/2} \left\| \widehat{\Gamma}_k^{-1} - \Gamma_k^{-1} \right\|_1 \xrightarrow{p} 0$ and $E \left(\left\| k^{-1/2} (T - k - h)^{1/2} U_{1T} \right\| \right) \leq s (T - k - h)^{1/2} \sum_{j=k+1}^{\infty} \|A_j^h\| \xrightarrow{p} 0$; $k, T \rightarrow \infty$ 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

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

and substituting this result into their proof. ■

Proof. Theorem 3

Follows directly from Lewis and Reinsel (1985) by redefining

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

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

Proof. Theorem 4

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

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

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

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

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 $\mathbf{f}(\cdot)$. For this reason, we directly assume here that stochastic conti-

⁵ Stochastic equicontinuity: For every $\epsilon, \eta > 0$ there exists a sequence of random variables $\widehat{\Delta}_t$ and a sample size t_0 such that for $t \geq t_0$, $\text{Prob}(|\widehat{\Delta}_T| > \epsilon) < \eta$ and for each ϕ there is an open set N containing ϕ with $\sup_{\tilde{\phi} \in N} \left| \widehat{Q}_T(\tilde{\phi}) - \widehat{Q}_T(\phi) \right| \leq \widehat{\Delta}_T$, for $t \geq t_0$.

nity holds and we refer the reader to Andrews (1994a, b) 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

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

which are

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

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

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

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

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

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

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

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

$$\begin{aligned} F_\phi(\hat{b}_T; \bar{\phi}) &\xrightarrow{p} F_\phi(b_0; \phi_0) = F_\phi \\ F_b(\bar{b}; \phi_0) &\xrightarrow{p} F_b(b_0; \phi_0) = F_b \end{aligned}$$

and hence

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

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

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

Since we know that

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

then

$$\sqrt{T}(\hat{\phi}_T - \phi_0) \xrightarrow{d} -(F'_\phi W F_\phi)^{-1} (F'_\phi W F_b) \sqrt{T}(\hat{b}_T - b_0)$$

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

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

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

$$\Omega_\phi = (F'_\phi W F_\phi)^{-1}$$

$$W = (F_b \Omega_b F'_b)^{-1}$$

■

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 $AR(\infty)$ Processes with Conditional Heteroskedasticity," *Econometric Reviews*, forthcoming.
- Gouriéroux, Christian and Alain Monfort (1997) **Simulation-Based Econometric Methods**. Oxford, U.K.: Oxford University Press.
- Hurvich, Clifford M. and Chih-Ling Tsai (1989) "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76(2): 297-307.
- Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1): 161-182.
- Lewis, R. A. and Gregory C. Reinsel (1985) "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16(33): 393-411.
- 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.
- 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)

		$\pi_1 = 0.25 \quad \theta_1 = 0.5$						T = 50
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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								
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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								
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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	0.25	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)

		$\pi_1 = 0.5 \quad \Theta_1 = 0.25$						T = 50
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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								
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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								
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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			
		<i>h = 2</i>		<i>h = 5</i>		<i>h = 10</i>	
		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							
		<i>h = 2</i>		<i>h = 5</i>		<i>h = 10</i>	
		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							
		<i>h = 2</i>		<i>h = 5</i>		<i>h = 10</i>	
		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					
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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									
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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	-0.15	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									
		<i>h</i> = 2		<i>h</i> = 5		<i>h</i> = 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 – 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 ($h = 27$)	HP	0.48 (0.017)	-0.0142 (0.0052)
PMD ($h = 11$)	ST	0.45 (0.034)	-0.030 (0.013)

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.

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

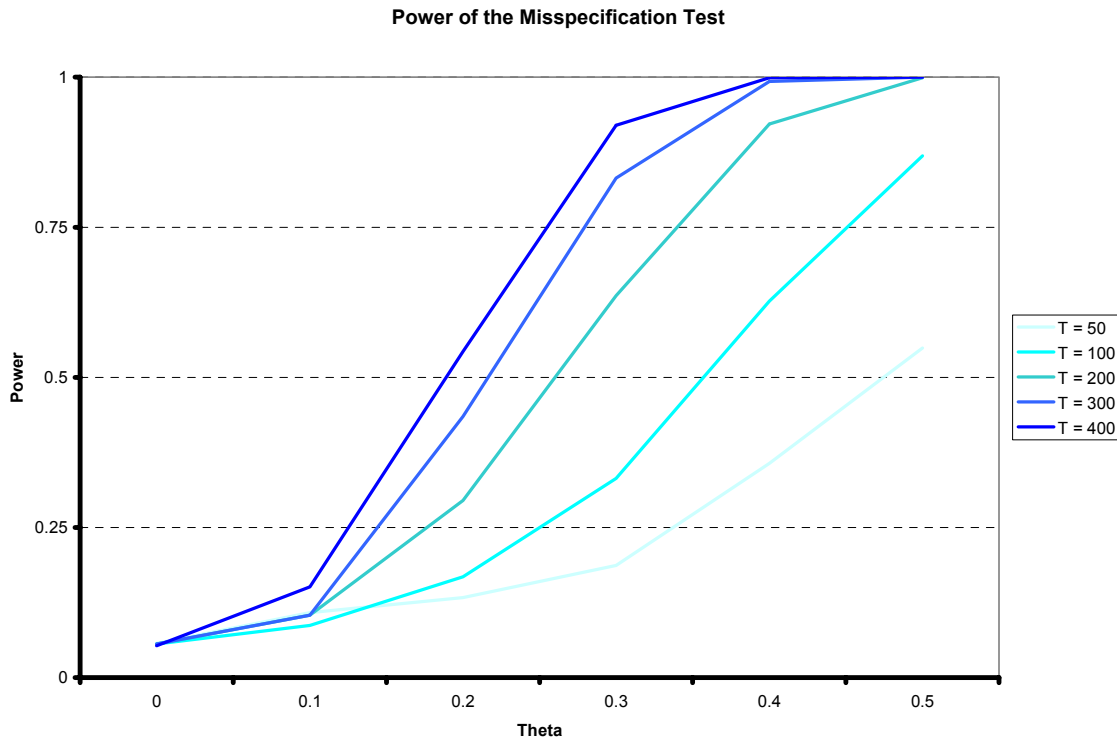
Estimates of Inflation Euler Equation: 1966:Q1 to 2001:Q4

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

Method	Specification	μ (S.E.)	γ (S.E.)
GMM	HP	0.66 (0.13)	-0.055 (0.072)
GMM	ST	0.63 (0.13)	-0.030 (0.050)
GMM	RULC	0.60 (0.086)	0.053 (0.038)
MLE	HP	0.17 (0.037)	0.10 (0.042)
MLE	ST	0.18 (0.036)	0.074 (0.034)
MLE	RULC	0.47 (0.024)	0.050 (0.0081)
OI-GMM	HP	0.23 (0.093)	0.12 (0.042)
OI-GMM	ST	0.21 (0.11)	0.097 (0.039)
OI-GMM	RULC	0.45 (0.028)	0.054 (0.0081)
PMD ($h = 10$)	HP	0.67 (0.06)	-0.036 (0.027)
PMD ($h = 11$)	ST	0.64 (0.06)	-0.029 (0.019)
PMD ($h = 11$)	RULC	0.67 (0.07)	0.033 (0.020)

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.

Figure 1 - Power of Misspecification Test



Notes: We simulate data from the model

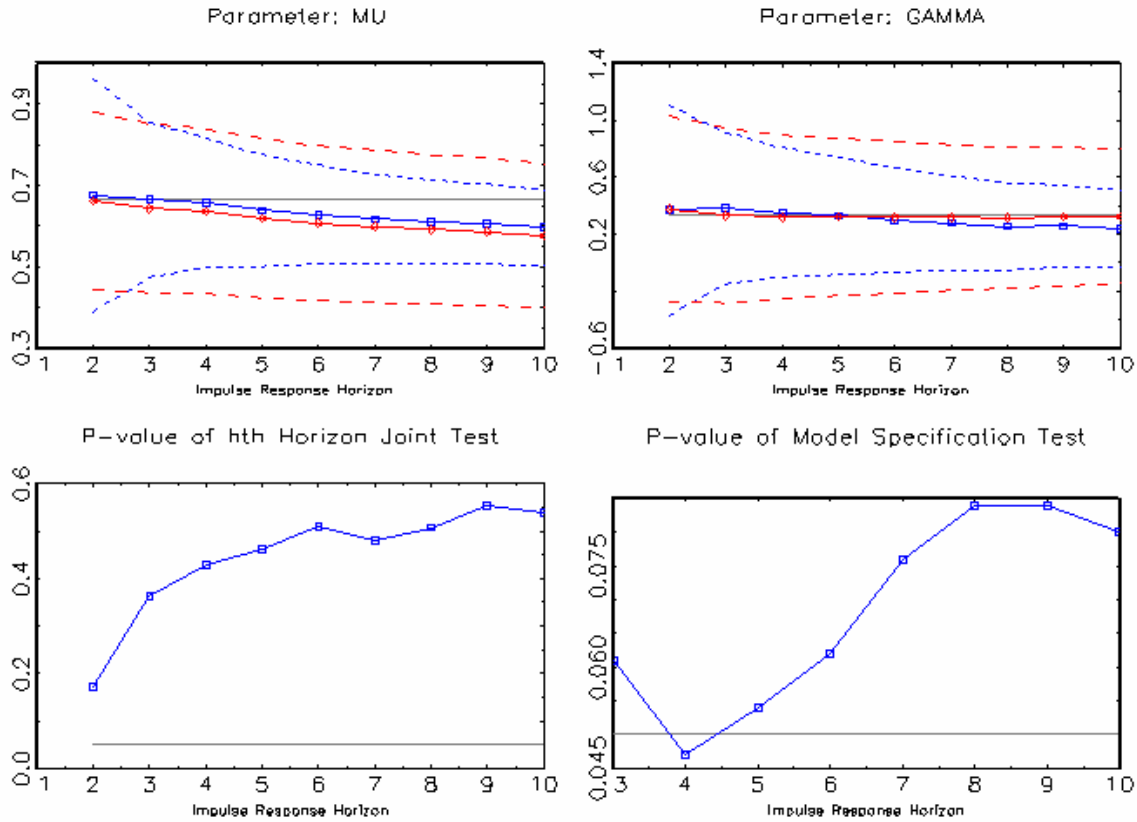
$$y_t = 0.5y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \quad \varepsilon_t \approx N(0,1)$$

where $\theta = 0; 0.1; 0.2; 0.3; 0.4; \text{ and } 0.5$. Sample sizes are $T = 50, 100, 200, 300, \text{ and } 400$ with 500 initial burn-in observations. Given the generated data, we estimate the model

$$y_t = \rho y_{t-1} + u_t$$

by projection minimum distance with impulse response horizon 5, and then calculate the chi-square test of misspecification. The experiment is repeated for 1,000 Monte Carlo replications and the rejection frequency of the chi-square test at a conventional 95% confidence level is reported.

Figure 2 – PMD and GMM Comparison when the Euler Equation is Correctly Specified. Sample Size = 100

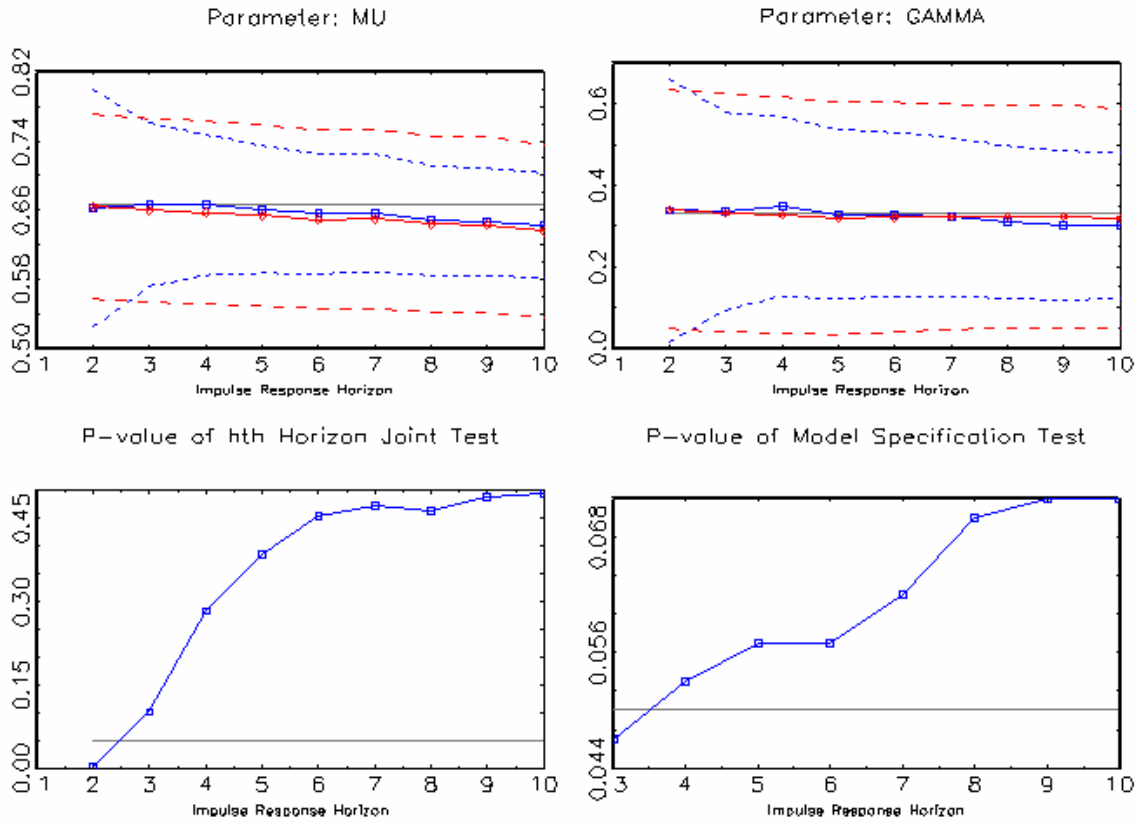


Notes: The top two panels display the Monte Carlo averages of the parameter estimates and associated two standard error bands. The bottom left panel reports the average p-value of the joint significance test on the coefficients of the h^{th} horizon whereas the bottom right panel is the power of the 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 3 – PMD and GMM Comparison when the Euler Equation is Correctly Specified. Sample Size = 400

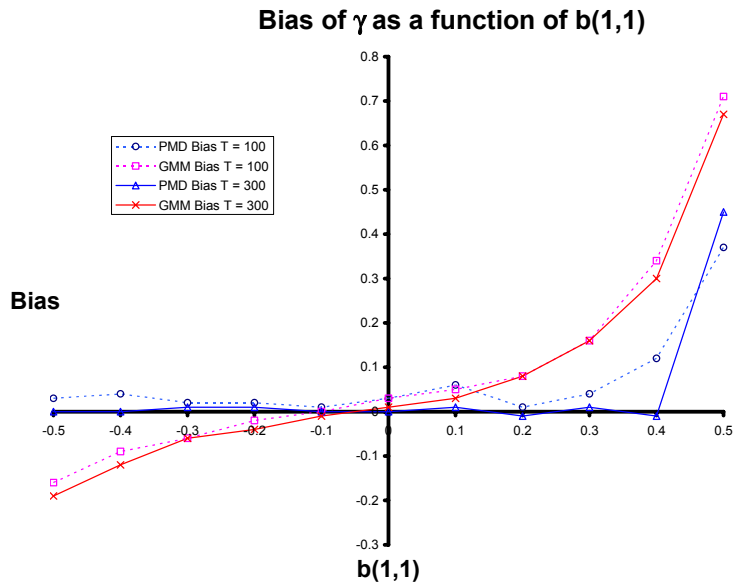
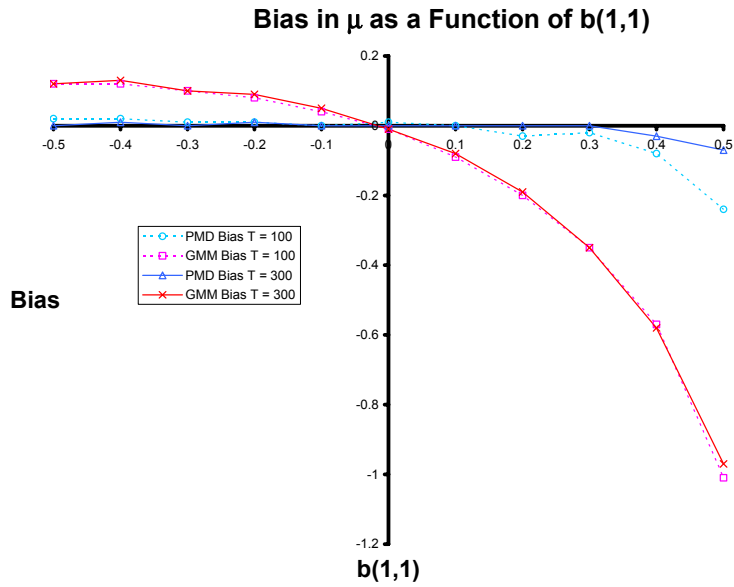


Notes: The top two panels display the Monte Carlo averages of the parameter estimates and associated two standard error bands. The bottom left panel reports the average p-value of the joint significance test on the coefficients of the h^{th} horizon whereas the bottom right panel is the power of the 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 :

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

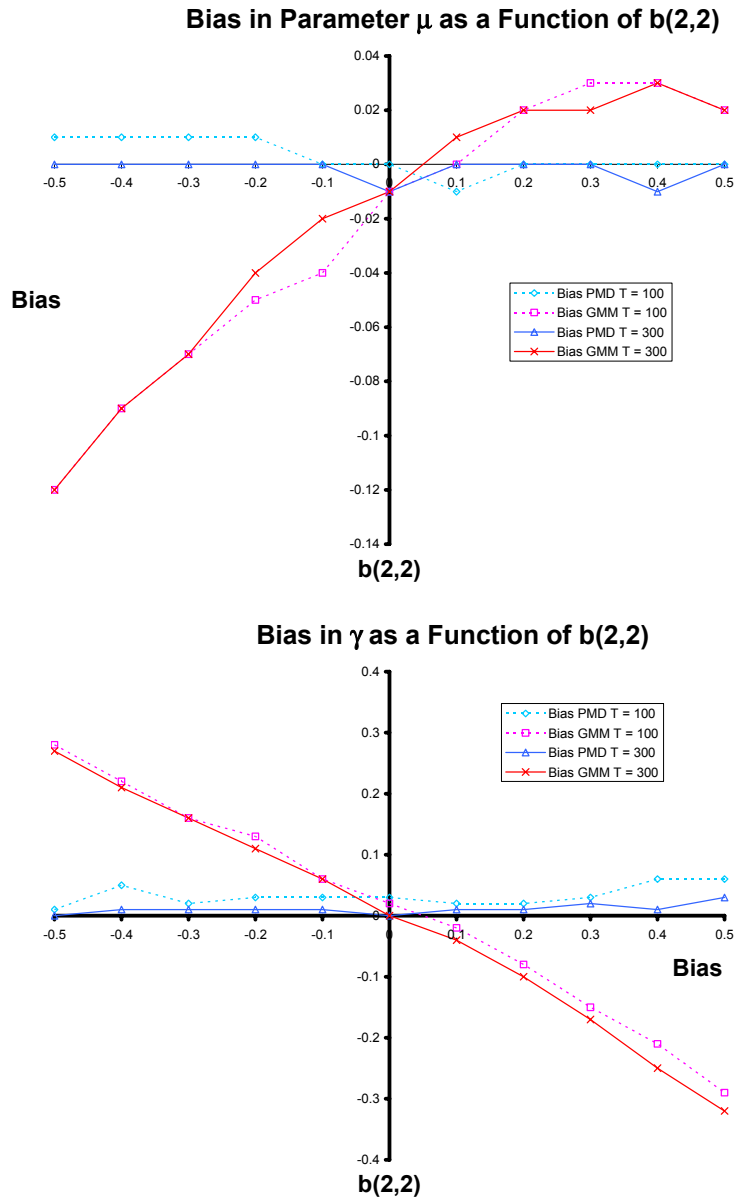
which implies that $\mu = \frac{2}{3}; \gamma = \frac{1}{3}$.

Figure 4 – 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.

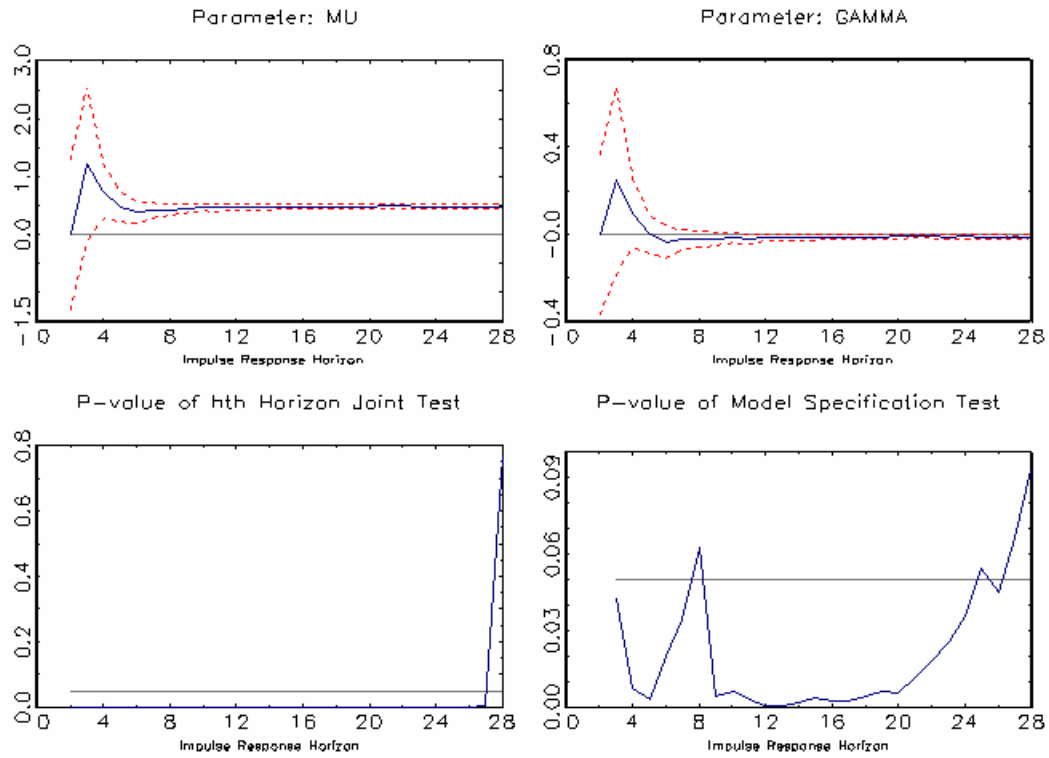
Figure 5 – PMD vs. GMM: Biases Generated by Neglected Dynamics in the Exogenous Variable



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 6 – Output Euler PMD Parameter Estimates

HP Detrended Output



Segmented Trend Output

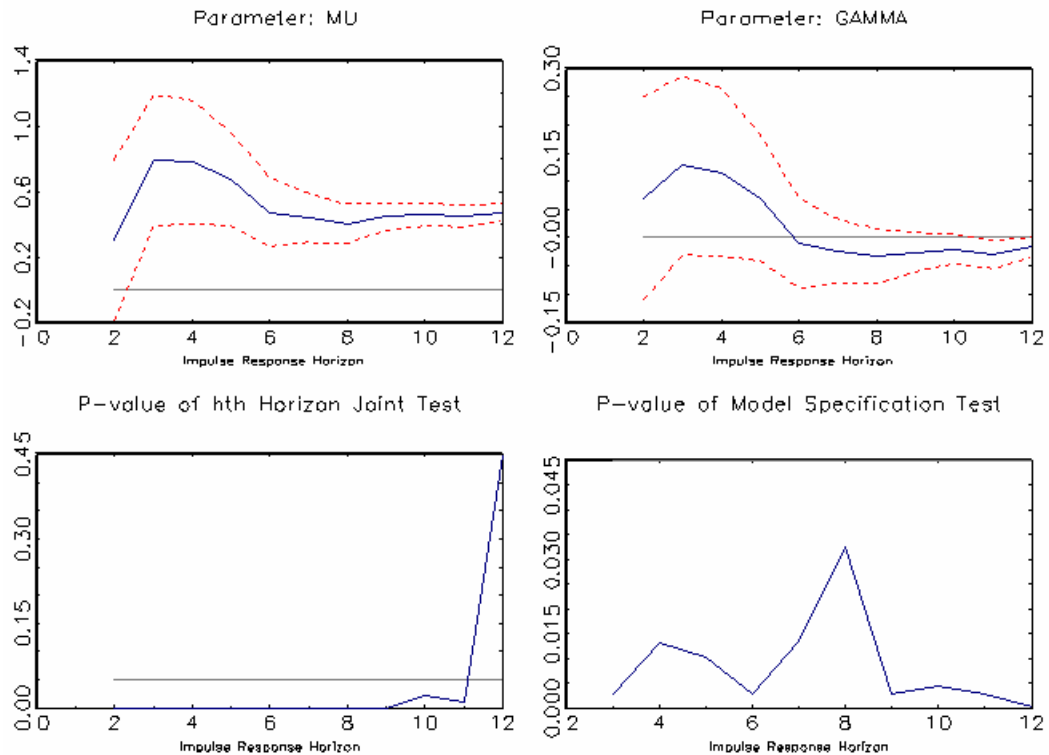
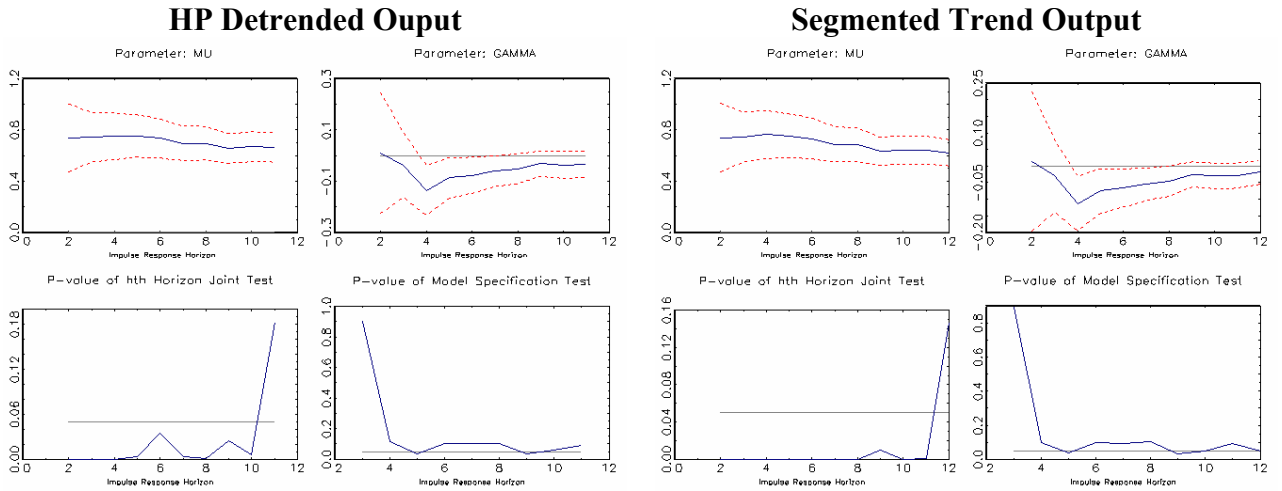


Figure 7 – Estimates of Inflation Euler Equation



Real Unit Labor Costs

