This version, September 2009

Estimation and Inference by the Method of Projection Minimum Distance: An Application to the New Keynesian Hybrid Phillips Curve*

Abstract

In most macroeconomic models, the stability of the solution path implies that the system is covariance-stationary and hence admits a Wold representation. The ability to estimate this Wold representation semi-parametrically by local projections (Jordà, 2005), even when the process for the solution path is unknown or unconventional, can be exploited to estimate the model's parameters by minimum distance techniques. We label this two-step estimation procedure "projection minimum distance" (PMD) and formally investigate its statistical properties in models where the mapping between Wold coefficients and parameters is linear even though the likelihood score function is nonlinear in the parameters, which traditionally requires numerical routines to maximize the likelihood. As an illustration of the practicalities of PMD estimation, we reexamine estimates of the New Keynesian Hybrid Phillips curve by providing ample Monte Carlo evidence and an empirical reassessment of Fuhrer and Olivei (2005).

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

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

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

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Bank of Canada. We thank Colin Cameron, Timothy Cogley, David DeJong, Richard Dennis, Stephen Donald, David Drukker, Jeffrey Fuhrer, James Hamilton, Peter Hansen, Kevin Hoover, Giovanni Olivei, Peter Robinson, Paul Ruud, Frank Schorfheide, Aaron Smith, Harald Uhlig, Frank Wolak and seminar participants at the Bank of Italy, Bocconi Univeristy - IGIER, Duke University, the Federal Reserve Bank of Dallas, the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of New York, Federal Reserve Bank of San Francisco, Federal Reserve Bank of St. Louis, Southern Methodist University, Stanford University, the University of California, Berkeley, the University of California, Davis, the University of California, Riverside, the University of Houston, University of Kansas, the University of Pennsylavania, the University of Texas at Austin, and the 2006 Winter Meetings of the American Economic Association in Boston, the 2006 European Meetings of the Econometric Society in Vienna, and the 3^{rd} Macro Workshop in Vienna, 2006 for useful comments and suggestions. Jordà is thankful for the hospitality of the Federal Reserve Bank of San Francisco during the preparation of this paper.

1 Introduction

Econometric estimation of dynamic stochastic general equilibrium (DSGE) models confronts model tractability against the wealth of dynamic interactions observed in reality. Approaches based on model implied likelihoods, be it through the classical (e.g. Sargent, 1989; and more recently Canova, 2007) or through the Bayesian (e.g. DeJong, Ingram and Whiteman, 2000; and An and Schorfheide, 2007) approaches, are only sensible with sufficiently complex models that narrow this separation with reality, and when suitable sources of exogenous variation are properly ascertained. The inability to conduct controlled experiments in macroeconomics and the capriciousness of natural or quasi-natural experiments often limit a practitioner's choice to estimation strategies based on heroic structural assumptions or appropriate instrumental variables techniques.

This paper introduces a statistical method of parameter estimation in which the economic model's restrictions are cast against a flexible, semi-parametric representation of the data based on its Wold (or impulse response) representation. The estimation methodology is particularly well-suited for models designed to capture dynamic co-movement, such as real business cycle models or New-Keynesian specifications, whose performance is often evaluated on the ability to match time series properties of macroeconomic data. The objective is to obtain parameter estimates even when the behavioral model offers incomplete characterizations of the dynamics and/or the forcing variables.

The focus of our paper is on specifications with likelihood score functions that are nonlinear in the parameters but where the mapping between Wold coefficients and parameters is linear. This may appear restrictive but in practice it is not: many models, such as linear or linearized Euler equations in rational expectations models (such as the Phillips curve examples we investigate in sections 4 and 5) and the more traditional ARMA(p,q) time series class (see section 4 for a Monte Carlo example) possess this feature. PMD provides estimates for this type of problems that are computationally simple (based on two least-squares steps) and whose large sample properties we derive formally in section 3.

Our estimator belongs to the class of limited-information, minimum-distance estimators, such as generalized method of moments (GMM); as well as other estimators popular in macroeconomic applications, such as Sbordone's (2002) forecast matching estimator (see Cogley and Sbordone, 2008, for a recent example), and Rotemberg and Woodford's (1997) impulse response matching estimator (see Christiano, Eichenbaum and Evans, 2005, for a recent application). However, formal statistical analysis on these last two estimators is largely unavailable, with little or no discussion about large sample results of estimated parameter covariance matrices, which of course are important for formal hypothesis testing and model evaluation.

We frame our discussion in the context of the voluminous literature that investigates inflation dynamics (e.g. volume 52 of the Journal of Monetary Economics in 2005). A critical divide in this literature appears to emerge between proponents of limited information, single-equation, instrumental-variable based methods (primarily in that issue Galí, Gertler and López-Salido, 2005) versus its critics and proponents of full information methods (e.g. Kurmann, 2005; Lindé, 2005; Rudd and Whelan, 2005), where a complete, New Keynesian formulation of the economy is often required. Central to this line of research is the desire to determine from the data the degree of backward/forward looking behavior of the Phillips curve because it is so central in determining optimal monetary policy responses, sacrifice ratios, and the stability of competing policy prescriptions (see e.g., Levin and Williams, 2003).

For this reason, we present our methods using examples based on the Phillips curve (see the extensive Monte Carlo experiments reported in section 4) although our methods are not limited to the examples that we provide. Moreover, our paper does not explicitly investigate situations in which the mapping between Wold coefficients and parameters is nonlinear but these are clearly of interest and are left for future research. Along the way, our paper provides some important intermediate results, such as deriving asymptotic approximations for local projection estimates (Jordà, 2005 and 2009) of impulse responses for a broad class of covariance-stationary processes that are not available elsewhere (see section 3). In section 5 we provide an empirical application of our methods to the dataset in Fuhrer and Olivei (2005), where we estimate IS and Phillips curves common in New Keynesian models. Concluding remarks are offered in section 6.

2 Projection Minimum Distance

The dynamics of many macroeconomic models often depend on expectations about their future values. In such specifications, the relative significance of forward versus backward looking terms is of considerable importance in determining optimal policy responses – the stability of the solution paths and the economy often depend on this feature. Unfortunately, because expectations are based on the same information set that determines backward-looking behavior, it is empirically difficult to disentangle which type of behavior is dominant. Single-equation, limited-information estimation methods therefore require appropriate instrumental variables, while full-information approaches based on the likelihood (classical or Bayesian) require complete and correctly specified models of the economy that describe how available information is allocated.

This section introduces our new estimation approach, PMD. We showcase the basic elements of the method with a completely stripped-down model that is nevertheless familiar to macroeconomists and easily scalable. Section 3 provides a more general treatment of PMD and formally derives its statistical properties. Consider the problem of estimating an Euler expression such as

$$y_t = \gamma E_t y_{t+1} + (1-\gamma) y_{t-1} + \varepsilon_t \tag{1}$$

where y_t and ε_t are scalar and ε_t is i.i.d. with finite first and second moments. Expression (1) contains forward- and backward-looking elements but no forcing variables, although scaling this example to include vector processes (as we do in section 3) would be sufficient to describe a large class of common specifications of system Euler equations such as the specifications common in the DSGE literature.

The stable solution path for this process in first order form is

$$y_t = \beta y_{t-1} + \delta \varepsilon_t \tag{2}$$

with

$$\left.\begin{array}{l} \beta = \frac{1-\gamma}{\gamma} \\ \delta = \frac{1}{\gamma} \end{array}\right\} \text{ for } \frac{1}{2} < \gamma \leq 1. \tag{3}$$

Thus, the maximum likelihood estimator (MLE) of expression (1) consists in estimating (2) given (3), that is

$$y_t = \frac{1-\gamma}{\gamma} y_{t-1} + \frac{1}{\gamma} \varepsilon_t \tag{4}$$

assuming, say, that the ε_t are Gaussian and possibly imposing the restriction $1/2 < \gamma \leq 1$, and using numerical routines to maximize the likelihood since (4) is nonlinear in the parameter γ. Alternatively, one could choose a prior for γ, and obtain the posterior density with the likelihood implied by (4) using an importance sampler that restricts draws of γ to be in the interval $(1/2, 1]$ and with Markov Chain Monte Carlo (MCMC) simulation techniques.

The stability of the solution in expression (2) implies that y_t has a Wold decomposition given by

$$y_t = \sum_{h=0}^{\infty} b_h v_{t-h} \tag{5}$$

with the convention $b_0 = 1$, and $\{v_t\}$ an i.i.d. sequence with mean zero and finite second order moments. Notice that $b_h = \beta^h$ and $v_t = \delta \varepsilon_t$ for β and δ in expression (3), making clear that (5) is an expression of y_t in terms of its reduced-form residuals.

In fact, substituting expression (5) into expression (2) and matching terms, results in the following set of conditions

$$b_h = \gamma b_{h+1} + (1-\gamma) b_{h-1} \text{ for } h \geq 1 \tag{6}$$

or rearranging terms further

$$(b_h - b_{h-1}) = \gamma (b_{h+1} - b_{h-1}) \text{ for } h \geq 1.$$

PMD consists in exploiting expression (6) using a first stage, semi-parametric estimate of b_h with local projections (Jordà, 2005 and 2009) and then using a second stage minimum distance step (Ferguson, 1958) to estimate γ. Specifically, recursive substitution of expression (2) results in

$$y_{t+h} = b_h y_t + b_{h-1} v_{t+1} + ... + b_1 v_{t+h-1} + v_{t+h} \tag{7}$$

where, $b_h = \beta^h$ in this simple example. From expression (7), a consistent estimate of b_h is simply

$$\widehat{b}_h = \frac{\sum_{t=1}^{T-h} y_{t+h} y_t}{\sum_{t=1}^{T-h} y_t^2}. \tag{8}$$

Although expression (6) provides in principle, an infinite number of conditions relating the Wold coefficients b_h and the parameter γ, we only need one condition to obtain an estimate $\widehat{\gamma}$. In this simple example all conditions for $h \geq 2$ are redundant since $b_h = \beta b_{h-1}$.

A minimum distance estimator of γ can be easily obtained by combining (8) with (6):

$$\widehat{\gamma}_{PMD} = \frac{\widehat{b}_2 - 1}{\widehat{b}_1 - 1} = \frac{\sum_{t=2}^{T-1} y_{t+1} y_{t-1} - \sum_{t=2}^{T-1} y_{t-1}^2}{\sum_{t=2}^{T-1} y_t y_{t-1} - \sum_{t=2}^{T-1} y_{t-1}^2} = \frac{\frac{1}{T-1} \sum_{t=2}^{T-1} (y_{t+1} - y_{t-1}) y_{t-1}}{\frac{1}{T-1} \sum_{t=2}^{T-1} (y_t - y_{t-1}) y_{t-1}} = \widehat{\gamma}_{GMM}. \tag{9}$$

where, in this example, PMD is equivalent to estimating (1) by GMM using y_{t-1} as an instrument.

In practice, the dynamics of the data rarely can be so concisely summarized with a first-order process. Similarly, the reason for using local projections is to avoid having to make assumptions about the nature of the reduced form process. Consider now a slight generalization to the set-up in expression (1), with shocks are that are serially correlated as is common in macroeconomic DSGE specifications:

$$\begin{cases} y_t = \gamma E_t y_{t+1} + (1-\gamma) y_{t-1} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \text{ for } |\rho| < 1 \end{cases}. \tag{10}$$

It is no longer sufficient to conjecture a first-order solution as in expression (2) and instead, the solution will be second-order autoregressive, say

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \delta \varepsilon_t.$$

Accordingly, recursive substitution reveals that

$$y_{t+h} = b_h y_t + c_h y_{t-1} + b_{h-1} v_{t+1} + ... + b_1 v_{t+h-1} + v_{t+h} \tag{11}$$

and therefore, the local projection estimator of b_h is not that in expression (8) but instead

$$\widehat{b}_h = \frac{\sum_{t=1}^{T-h} y_{t+h} M_{t-1} y_t}{\sum_{t=1}^{T-h} y_t M_{t-1} y_t}$$

where M_{t-1} simply projects the effect of y_{t-1} in expression (11) as in Jordà's (2005) original formulation (since the coefficient c_h is of no practical interest). Proceeding as in expression (9), the minimum distance estimator of γ is now

$$\widehat{\gamma}_{PMD} = \frac{\frac{1}{T-2}\sum_{t=3}^{T-1}(y_{t+1} - y_{t-1})M_{t-2}y_{t-1}}{\frac{1}{T-2}\sum_{t=3}^{T-1}(y_t - y_{t-1})M_{t-2}y_{t-1}} \neq \frac{\frac{1}{T-2}\sum_{t=3}^{T-1}(y_{t+1} - y_{t-1})y_{t-1}}{\frac{1}{T-2}\sum_{t=3}^{T-1}(y_t - y_{t-1})y_{t-1}} = \widehat{\gamma}_{GMM} \qquad (12)$$

Serial correlation in the shocks means that y_{t-1} is no longer a legitimate instrument for GMM estimation since $E(u_t|y_{t-1}) \neq 0$. However, $\widehat{\gamma}_{PMD}$ is consistent for γ because the term M_{t-2} in expression (12) effectively orthogonalizes the moment conditions against the omitted information dated at time $t-2$, which is the source of the GMM bias.

Expression (12) highlights a larger point that has, perhaps, received insufficient attention in the literature (some exceptions are Jondeau and Le Bihan, 2008; Mavroeidis, 2004; and Rudd and Whelan, 2006): the validity of the instruments in GMM estimation is determined not by one's hypothesized model (in our case, expression (1)) but by the true data generating process (DGP), in the example, expression (10). Since tractable macroeconomic models require parsimonious specifications in terms of dynamics and driving variables, it is important to assess the validity of candidate instruments for GMM estimation.[1] A natural solution is to orthogonalize candidate instruments with respect to potentially omitted information (be it in the form of omitted dynamics, as in the example, and/or omitted driving variables).

Local projection estimates b_h semi-parametrically and therefore takes on a broader perspective about the underlying DGP that allows some of these biases to be naturally corrected (as in expression (12)). In practical settings, it therefore seems desirable that local projections be estimated from systems that not only include the variables described by the theoretical model but include variables that are simply deemed to be good predictors.

[1] MLE will be inconsistent due to misspecification although the magnitude of the bias will depend from one application to the other.

The preceding discussion is useful to highlight other important features of our new estimator. Specifically, we have emphasized that the Wold representation in expression (5) used to derive the mapping between Wold coefficients and structural parameters, is in terms of reduced-form forecast errors. The reason is that, to the extent that the parameters of interest can be identified from second order properties of the data, one is not required to recover the structural impulse response representation. Formulation of PMD in terms of the structural representation does not complicate the derivation of its statistical properties but opens the door to a number of important potential problems. First, although some advances have been made recently (see e.g. Granger and Swanson, 1997; and Demiralp and Hoover, 2003), structural identification largely relies on statistically untestable assumptions (such as the common recursive zero restrictions embodied by Cholesky factorizations of the residual covariance matrix). Second, as Fernández-Villaverde, Rubio-Ramírez, Sargent and Watson (2007) have pointed out, there is a class of macroeconomic models whose structure cannot be recovered by recursive identification assumptions, an observation that would introduce unwanted biases in PMD estimation.

In that same paper, Fernández-Villaverde et al. (2007) discuss how the solution to several popular macroeconomic models is of a VARMA form rather than the customary VAR. Here too PMD is advantageous relative to other methods since, as we will show in section 3, local projection estimates of the Wold coefficients are consistently estimated for a large class of primitive time series representations that include most ARMA-type models. In fact, PMD can be used to estimate an ARMA model conveniently because the mapping between Wold coefficients and parameters is linear. For instance, suppose the DGP is

$$y_t = \rho y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \tag{13}$$

then plugging-in the Wold representation in expression (5) and equating terms, we have

$$b_h = \rho b_{h-1} + \theta d_h \text{ for } h \geq 1 \tag{14}$$

where $d_h = 1$ for $h = 1$ and is zero otherwise. Thus, least-squares projection of $\widehat{b}_h$ onto $\widehat{b}_{h-1}$ and d_h

for $h = 1, ..., H$ with H some truncation horizon, provides a simple computational method to obtain estimates of ρ and θ. In fact as we show in a Monte Carlo exercise in section 4, PMD estimates of these parameters are consistently and very efficiently estimated relative to the maximum likelihood Cramer-Rao lower bound even in relatively small samples.

Linear Euler equations and ARMA models are two examples where the mapping between Wold coefficients and parameters is linear whereas the likelihood score function is nonlinear in the parameters, requiring numerical or simulation techniques for its maximization. Further, in rational expectations models, there is often more than one possible solution so that choices have to be made before the likelihood can be fully specified. In contrast, PMD only requires two rather straight-forward least-squares steps, thus making the method computationally attractive and numerically robust.

Naturally, an important question is to evaluate the relative efficiency of PMD estimates. In part, we broach this issue in the Monte Carlo simulations in section 4. Here we discuss a few practical observations. The Wold decomposition is an infinite orthogonal decomposition of a time series so that, in principle, it provides an infinite set of conditions with which to estimate the parameters of a model. Several factors advise for a more parsimonious approach instead. The example motivated by expression (1) suggests that only the moment condition associated with $h = 1$ in expression (6) is needed since $b_h = \beta^h$ and conditions for $h \geq 2$ do not provide independent sources of variation to identify the parameter γ.

More generally, the DGP is unknown so it is possible to contemplate using numerous moment conditions. However, the covariance-stationarity requirement implied by the stability of the solution paths suggests that the b_h tend to zero at, possibly, an exponential rate of decay as h grows. In the example of expression (6), recall that $b_h = \beta^h$ with $|\beta| < 1$. For example, if $\beta = 0.5$, then $b_5 = 0.03125$, $b_6 = 0.015625$, and $b_7 = 0.0078125$, numbers that are considerably smaller than reasonable estimation error in the type of finite samples available, and therefore, more likely to introduce additional small sample error than identification variability. This problem is not dissimilar to the weak instrument problem in GMM (see e.g. Stock, Wright and Yogo, 2002) and suggests one must carefully choose the truncation parameter H. Fortunately, Hall, Inoue, Nason and Rossi (2007) provide a simple information criterion for selecting the optimal horizon H, specifically

$$\widehat{H} = \arg \min_{H \in \{h_{\min}, \ldots, h_{\max}\}} \ln\left(\left|\widehat{\Omega}_{\gamma}\right|\right) + h \frac{\ln\left(\sqrt{T/k}\right)}{\left(\sqrt{T/k}\right)} \tag{15}$$

where $h_{\min}$ is selected so that there are enough conditions to at least have just identification of structural parameter estimates. The next section presents the statistical properties of our method formally.

3 Statistical Properties of PMD

This section derives the large sample properties of our estimator in a general setting that includes vector processes. The estimator consists of two stages. In the first stage we estimate the first H terms of the Wold representation semi-parametrically with local projections and for this reason, we will first derive the asymptotic properties of local projections in general settings. The second stage of our estimator consists in minimizing the distance resulting from a linear mapping between the Wold coefficients and the coefficients of the model we are interested in. Therefore, we will provide asymptotic results based on this mapping and the first-stage asymptotic results for local projections. As a way to organize the discussion with general notation for a wide range of models, consider the covariance-stationary, $n \times 1$ vector $\mathbf{y}_t$ with Wold representation given by

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{h=0}^{\infty} B_h \mathbf{u}_{t-h} \tag{16}$$

where the $\mathbf{u}_t$ are *i.i.d.* mean zero and with finite covariance matrix. Next, let $\boldsymbol{\gamma}$ be an $nr \times 1$ vector of parameters of interest, say r regressors in n equations. We specify the mapping between the B_h and the $\boldsymbol{\gamma}$ as follows

$$S_y \mathbf{B}' = S_x (I_r \otimes \mathbf{B}') \boldsymbol{\gamma} \tag{17}$$

where the $nH \times n$ matrix $\mathbf{B}' = (B_1 \quad \ldots \quad B_H)'$ collects the $n \times n$, B_h matrices in (16) for some truncation horizon H that can be determined with the information criterion in expression (15).

The matrices S_y and $S_x = (S_1 \quad ... \quad S_r)$ are selector matrices that pick the appropriate elements in $\mathbf{B}'$ to describe the mapping of the particular model of interest. The use of the matrix notation in (17) is meant to simplify the task of programming our estimator with standard software packages.

As an example, suppose that we want to estimate the model

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times n}{\Gamma_1} \underset{n\times 1}{E_t\mathbf{y}_{t+1}} + \underset{n\times n}{\Gamma_2} \underset{n\times 1}{\mathbf{y}_{t-1}} + \underset{n\times 1}{\boldsymbol{\varepsilon}_t}$$

which generalizes the Euler example in expression (1) to an unrestricted vector system. Substituting (16) into this expression and equating terms, we have the following linear mapping

$$B_h = \Gamma_1 B_{h+1} + \Gamma_2 B_{h-1} \text{ for } h \geq 1. \tag{18}$$

In this example, (18) can be cast in the form (17) by noticing that the matrix S_y is the result of deleting the first and last n rows of the identity matrix I_{nH}; the matrix S_1 is the result of deleting the last $2n$ rows of the identity matrix I_{nH}; and S_2 is the result of deleting the first $2n$ rows of the identity matrix I_{nH}; with $S_x = (S_1 \quad S_2)$. Finally notice that $\boldsymbol{\gamma} = vec(\Gamma_1 \quad \Gamma_2)$.

We note that the mapping in (17) could have included $B_0 = I_n$ and therefore we would have had $\overline{\mathbf{B}}' \equiv (I_n \quad B_1...B_H)'$. If $\Omega_{\mathbf{b}}$ denotes the covariance matrix for $vec(\widehat{\mathbf{B}})$, then all that is required to construct $\overline{\Omega}_{\mathbf{b}}$, the covariance matrix of $vec(\widehat{\overline{\mathbf{B}}})$, is to extend the matrix $\Omega_{\mathbf{b}}$ with matrix $\mathbf{0}_n$ in the first $n \times n$ diagonal block, and zeroes in the other positions. Then substituting $\overline{\mathbf{B}}$ and $\overline{\Omega}_{\mathbf{b}}$ for $\mathbf{B}$ and $\Omega_{\mathbf{b}}$ (and estimators similarly) in the derivations that follow in sections 3.1, 3.2 and 3.3 would permit straight-forward extension of the estimating equations reported there. However, in the interest of keeping the exposition as simple as possible, we preferred the less notationally intensive approach that we follow below. We begin now by deriving the properties of the first-stage local projection estimator.

3.1 Statistical Properties of Local Projections: First Stage

Consider the $n \times 1$ covariance-stationary vector $\mathbf{y}_t$ with Wold representation given by (16) and where the $\mathbf{u}_t$ are *i.i.d.,* mean zero, homoscedastic and with finite covariance matrix Σ_u and the B_h

satisfy $\sum_{h=0}^{\infty} ||B_h|| < \infty$ where $||B_h||^2 = tr(B_h' B_h)$ with $B_0 = I_n$. Further, assume $\det\{B(z)\} \neq 0$ for $|z| \leq 1$ where $B(z) = \sum_{h=0}^{\infty} B_h z^h$ so that the process can be written in its infinite VAR representation

$$\mathbf{y}_t = \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \mathbf{u}_t$$

with $\sum_{j=1}^{\infty} ||A_j|| < \infty$ and $A(z) = B(z)^{-1}$. By recursive substitution

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

where $A_1^h = B_h$, $A_j^h = B_{h-1}A_j + A_{j+1}^{h-1}$ for $h \geq 1$ and with the conventions $A_{j+1}^0 = 0; B_0 = I_n$ with $j \geq 1$. Jordà's (2005) local projection estimator consists in estimating B_h from a least-squares estimate of A_1^h with the truncated regression

$$\begin{aligned} \mathbf{y}_{t+h} &= A_1^h \mathbf{y}_t + ... + A_k^h \mathbf{y}_{t-k+1} + \mathbf{v}_{k,t+h} \\ \mathbf{v}_{k,t+h} &= \sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} + \mathbf{u}_{t+h} + \sum_{j=1}^{h-1} B_j \mathbf{u}_{t+h-j} \end{aligned}$$

Proposition 1 ***Consistency.*** *Let $\{\mathbf{y}_t\}$ satisfy (16) and assume that: (i) $E|u_{it}, u_{jt}, u_{kt}, u_{lt}| < \infty$ for $1 \leq i, j, k, l \leq n$; (ii) k satisfies*

$$\frac{k^2}{T} \to 0; \quad T, k \to \infty;$$

(iii) k satisfies

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

Then

$$\widehat{A}_1^h \xrightarrow{p} B_h$$

The proof of the theorem is provided in the appendix and it is based on an extension to the arguments in Lewis and Reinsel (1985), who develop consistency and asymptotic normality results for truncated VAR estimates of infinite order processes. Notice that more general heteroskedasticity and mixing conditions on the $\mathbf{u}_t$ are possible but these extensions are beyond the scope of this paper and the reader is referred to Gonçalves and Kilian (2007) and Kuersteiner (2005) for related examples. Further notice that the assumption that the Wold representation is invertible restricts the scope of the method somewhat but includes a vast majority of common VAR, VARMA, and VMA representations. In practice (through Monte Carlo evidence), Jordà (2005) shows that the choice of truncation lag k has little impact on the consistency of the parameter matrix A_1^h, which is all that is needed in the minimum distance stage of the PMD estimator.

Proposition 2 ***Normality.*** *Let $\{\mathbf{y}_t\}$ satisfy (16) and assume that: (i) $E\left|u_{it}, u_{jt}, u_{kt}, u_{lt}\right| < \infty$ for $1 \leq i, j, k, l \leq n$; (ii) k satisfies*

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

(iii) k satisfies

$$\sqrt{T-k-H} \sum_{j=k+1}^{\infty} ||A_j|| \to 0; \quad T, k \to \infty.$$

Then

$$\sqrt{T-k-H}vec(\widehat{\mathbf{B}}_T - \mathbf{B}_0) \xrightarrow{d} N\left(0, \Omega_{\mathbf{b}}\right)$$

$$\Omega_{\mathbf{b}} = \left[(X'MX)^{-1} \otimes \Sigma_v\right]$$

$$\widehat{\Sigma}_v = \frac{\widehat{V}\widehat{V}'}{T-k-H}$$

where $\widehat{\mathbf{B}}_T = \left(X'MX\right)^{-1}\left(X'MY\right)$, Y is the $T \times nH$ matrix of observations for $(\mathbf{y}_{t+1}, ..., \mathbf{y}_{t+H})'$; X is the $T \times n$ matrix of observations for $\mathbf{y}_t$; $M = I - Z(Z'Z)^{-1}Z'$ where Z is the $T \times n(k+1)$ matrix of observations for $(1, \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-k+1})'$ and $\widehat{V} = MY - MX\widehat{\mathbf{B}}_T$. The proof is provided in the appendix. We remark that assumptions (ii) and (iii) provide an upper and a lower bound for the rate at which k has to grow with the sample size, respectively. For more intuition, Lütkepohl (2005, p. 533) provides a nice and simple example with a univariate MA(1) model.

3.2 Statistical Properties of Projection Minimum Distance: Second Stage

Given $\widehat{\mathbf{B}}_T$, consider estimating $\boldsymbol{\gamma}$ from the mapping (17) by minimizing

$$\min_{\boldsymbol{\gamma}} \widehat{Q}_T(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}) = f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right)' \widehat{W} f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right) \tag{20}$$

where $vec(\widehat{\mathbf{B}}_T) = \widehat{\mathbf{b}}_T$ and

$$f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right) = S_y \widehat{\mathbf{B}}_T' - S_x(I_k \otimes \widehat{\mathbf{B}}_T')\boldsymbol{\gamma}$$

where we use the notation $\widehat{\mathbf{b}}_T$ in $f(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma})$ instead of $\widehat{\mathbf{B}}_T$ for clarity and to be consistent with the derivations provided in the appendix. Let $Q_0(\boldsymbol{\gamma})$ denote the objective function at $\mathbf{B}_0$. Then the following lemma shows that the solution of this problem, $\widehat{\boldsymbol{\gamma}}_T$ is consistent for $\boldsymbol{\gamma}_0$.

Lemma 3 ***Consistency.*** *Given that* $\widehat{\mathbf{b}}_T \xrightarrow{p} \mathbf{b}_0$ *from proposition 1, assume that: (i)* $\widehat{W} \xrightarrow{p} W$ *is a positive semidefinite matrix; (ii)* $Q_0(\gamma)$ *is uniquely maximized at* $(\mathbf{B}_0, \boldsymbol{\gamma}_0) = \theta_0 \in \Theta$*; (iii) The parameter space* Θ *is compact; (iv)* $f(b_0, \gamma)$ *is continuous in a neighborhood of* $\gamma_0 \in \Theta$*; (v) instrument relevance condition:*

$$rank\,[WF_\gamma] = \dim(\boldsymbol{\gamma}) \text{ where } F_\gamma = \frac{\partial f(\mathbf{b}_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}};$$

(vi) identification condition: $\dim\left(f\left(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma}\right)\right) \geq \dim(\gamma)$.

Then

$$\widehat{\boldsymbol{\gamma}}_T \xrightarrow{p} \boldsymbol{\gamma}_0$$

The proof is provided in the appendix where it is worth remarking that the proof takes H to be finite and given. As we will see momentarily, condition (v) is easy to check since $F_\gamma = -S_x(I_r \otimes \widehat{\mathbf{B}}_T')$.

Lemma 4 ***Normality.*** *Assume: (i)* $\widehat{W} \xrightarrow{p} W$ *where* $W = (F_\mathbf{b}\Omega_\mathbf{b}F_\mathbf{b})^{-1}$*, a positive definite matrix and where* F_b *is defined as in assumption (v) below; (ii) Let* $\widehat{\mathbf{b}}_T \xrightarrow{p} b_0; \widehat{\boldsymbol{\gamma}}_T \xrightarrow{p} \gamma_0$ *from proposition 1 and lemma 3; (iii)* b_0 *and* γ_0 *are in the interior of* Θ*; (iv)* $f\left(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma}\right)$ *is continuously differentiable in a neighborhood* N *of* θ_0*; (v) There is a* F_b *and a* F_γ *that are continuous at* b_0 *and* γ_0 *respectively and*

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

(vi) For $F_\gamma = F_\gamma(\gamma_0)$ *then* $F_\gamma' W F_\gamma$ *is invertible.*

Then

$$\sqrt{T-H-k}\,(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) \xrightarrow{d} N\,(0, \Omega_\gamma)$$
$$\Omega_\gamma = \left(F_\gamma' W F_\gamma\right)^{-1}$$

The proof is provided in the appendix using the same principles required to derive the proof of asymptotic normality typical of GMM and minimum distance problems (see e.g., Newey and

McFadden, 1994; Wooldridge, 1994). We have taken the simpler route here of brushing aside weak instrument conditions/problems such as those discussed, e.g., in Bekker (1994), Staiger and Stock (1997), Stock, Wright and Yogo (2002) and many others with assumption (v) in Lemma 3, believing it was more useful to provide the foundational results first. Weak instrument problems that can arise with PMD are of a similar nature to those already investigated in the literature in a GMM context, so we refer the reader to this literature directly. In addition, we remark that it is probably wise to be parsimonious in the choice of H in light of the *many* (weak) instruments problem described in Andrews and Stock (2007).

In practice, we recommend choosing the optimal impulse response horizon using the information criterion in Hall et. al. (2007) in expression (15). In finite samples, all asymptotic expressions can be replaced by their usual small sample estimates. Lastly, note that the optimal weighting matrix $\widehat{W}_o = (F'_{\mathbf{b}}\Omega_{\mathbf{b}}F_{\mathbf{b}})^{-1}$ cannot be computed directly as $F_{\mathbf{b}}$ is a function of $\boldsymbol{\gamma}$. A consistent estimate of $\boldsymbol{\gamma}$ can be obtained with the equal-weights matrix $\widehat{W}_e = I$ (Lemma 3 only requires $\widehat{W}$ to be positive semidefinite to achieve consistency) and used to construct the optimal-weights $\widehat{W}_o$ and compute all the relevant statistics. In principle, one can iterate on this procedure to refine the estimates of $\boldsymbol{\gamma}$ although asymptotically, one iteration is sufficient.

Finally, lemma 4 and standard results are all that is needed to show that a test of overidentifying restrictions can be easily obtained by realizing that the minimum distance function $\widehat{Q}_T$ evaluated at the optimum $\widehat{\mathbf{B}}_T, \widehat{\boldsymbol{\gamma}}_T$ has a chi-square distribution with degrees of freedom $\dim\left(f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right)\right) - \dim(\boldsymbol{\gamma})$.

3.3 Summary of Practical Results

Suppose a model of interest can be set-up according to the mapping in expression (17). Then the PMD estimator consists of the following computations:

1. Compute the local projections estimates $\widehat{\mathbf{B}}_T$:

$$\widehat{\mathbf{B}}_T = (X'MX)^{-1}(X'MY)$$

with covariance matrix for $vec(\widehat{\mathbf{B}}_T) = \widehat{\mathbf{b}}_T$ given by

$$\widehat{\Omega}_{\mathbf{b}} = \left[(X'MX)^{-1} \otimes \widehat{\Sigma}_v \right]$$
$$\widehat{\Sigma}_v = \frac{\widehat{V}\widehat{V}'}{T-k-H}$$

where Y is the $T \times nH$ matrix of observations for $(\mathbf{y}_{t+1}, ..., \mathbf{y}_{t+H})'$; X is the $T \times n$ matrix of observations for $\mathbf{y}_t$; $M = I - Z(Z'Z)^{-1}Z'$ where Z is the $T \times n(k+1)$ matrix of observations for $(1, \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-k+1})'$ and $\widehat{V} = MY - MX\widehat{\mathbf{B}}_T$. The truncation lag k can be determined by an information criterion such as AIC, SIC, or preferably AIC$_c$, which is Hurvich and Tsai's (1989) small sample correction for AIC.

2. Compute the minimum distance estimate $\widehat{\boldsymbol{\gamma}}_T$ as

$$\widehat{\boldsymbol{\gamma}}_T = \left[(I_k \otimes \widehat{\mathbf{B}}_T) S_x' \widehat{W} S_x (I_k \otimes \widehat{\mathbf{B}}_T') \right]^{-1} \left[(I_k \otimes \widehat{\mathbf{B}}_T) S_x' \widehat{W} S_y \widehat{\mathbf{B}}_T' \right] \tag{21}$$

where the optimal weighting matrix based on $\widehat{W}_o = (F_{\mathbf{b}}' \widehat{\Omega}_{\mathbf{b}} F_{\mathbf{b}})^{-1}$ uses a first-stage estimate of $\boldsymbol{\gamma}$ based on expression (20) with $\widehat{W}_e = I$,

$$F_{\mathbf{b}} = \left[(I_n \otimes S_y) - (\widehat{\boldsymbol{\gamma}}_T \otimes S_x)(Q \otimes I_{nH}) \right] K;$$
$$Q = \begin{bmatrix} I_n \otimes e_1 \\ \vdots \\ I_n \otimes e_r \end{bmatrix};$$

and where e_j $j = 1, .., r$ are the columns of the identity matrix I_r and K the commutation matrix such that $vec(\mathbf{B}') = Kvec(\mathbf{B})$ as defined, for example, in Harville (1997). The way to see these results is to notice that

$$\begin{aligned} vec(f(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma})) &= vec(S_y \widehat{\mathbf{B}}_T) - vec(S_x (I_k \otimes \widehat{\mathbf{B}}_T) \boldsymbol{\gamma}) \\ &= (I_n \otimes S_y) K \widehat{\mathbf{b}}_T - (\widehat{\boldsymbol{\gamma}}_T \otimes S_x)(Q \otimes I_{nH}) K \widehat{\mathbf{b}}_T \end{aligned}$$

The truncation horizon H for the Wold representation coefficients can be chosen with Hall et al.'s (2007) information criterion in expression (15).

3. The optimal covariance matrix for $\widehat{\boldsymbol{\gamma}}_T$ is simply

$$\widehat{\Omega}_\gamma = (F'_\gamma \widehat{W} F_\gamma)^{-1}$$

where $\widehat{W}$ is the optimal weighting matrix given by $\widehat{W}_o = (F'_{\mathbf{b}}\widehat{\Omega}_{\mathbf{b}}F_{\mathbf{b}})^{-1}$ and where

$$F_\gamma = -S_x(I_k \otimes \widehat{\mathbf{B}}'_T)$$

Notice that the conditions for Lemmas 3 and 4 are easily verified for each candidate model based on expression (17) and the implied minimum distance function

$$f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right) = S_y \widehat{\mathbf{B}}'_T - S_x(I_k \otimes \widehat{\mathbf{B}}'_T)\boldsymbol{\gamma}.$$

4 Small-Sample Properties: Monte Carlo Experiments

This section investigates the small sample properties of PMD with two sets of experiments: one compares estimation of a conventional ARMA(1,1) by maximum likelihood with PMD. The other generates data from an extended version of the New Keynesian model in Lindé (2005) and compares the small sample properties of the New Keynesian Hybrid Phillips Curve estimated by GMM with PMD.

4.1 PMD vs. MLE

The data for this set of experiments is generated from the univariate ARMA(1,1) model

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

for the following four different pairs of parameter values: (1) $\rho = 0.25$, $\theta = 0.50$; (2) $\rho = 0.50$, $\theta = 0.25$; (3) $\rho = 0$, $\theta = 0.5$; and (4) $\rho = 0.5$, $\theta = 0$. The last two cases are a pure MA(1) and a pure AR(1) models but they will be specified as ARMA(1,1) models in the estimation.

Each of the 1,000 simulation runs has the following features. We use 500 burn-in replications to avoid initialization issues with sample sizes $T = 50$, 100, and 400. The lag length of the local projection step is determined automatically by AIC_C. For the minimum distance step, we experiment with fixed values $H = 2$, and 5. In an earlier version we also experimented with $H = 10$ to investigate if including many uninformative conditions would distort the estimates. Except for one case when $T = 50$ we did not find significant differences and hence the results are not reported here although they are available upon request. For $H = 2$, we have just-identification, otherwise, we have overidentifying restrictions. Given our choices of ρ and θ in all four cases, the impulse response coefficients are all very close to 0 for $H = 5$. In contrast to MLE, which requires numerical optimization routines, PMD for this example requires two very simple and numerically robust least-squares steps.

Table 1 reports Monte Carlo averages and standard errors of the parameter estimates calculated with the analytic formulas of the large-sample approximations. PMD estimates converge to the true parameter values quickly as the sample size grows even when starting from small sample sizes. It is surprising that including some conditions with little information value ($H = 5$ rather than $H = 2$) does not appear to distort the estimates (or the standard errors) with sample sizes as low as $T = 100$ observations. For sample sizes $T = 100$ and 400, PMD and MLE standard errors are virtually the same and comparable to the empirical Monte Carlo values (not reported here but available upon request). Finally, we remark that we had difficulty getting convergence of the MLE estimator for all the runs when the DGP was a pure MA(1) or AR(1) model. Rather than redo (or disregard) specific runs, we preferred to leave the results blank to highlight that although MLE ran into numerical difficulties, PMD was numerically stable and robust in all the cases. Overall, these experiments suggests that PMD has very good small sample properties, converging quickly to the theoretical values and with relatively the same efficiency as MLE even though PMD uses simple least squares algebra and MLE requires numerical routines to maximize the likelihood.

4.2 PMD vs. GMM

This set of experiments borrows several elements from the simulation study in Lindé (2005), which compares the small sample properties of GMM vs. FIML estimation of the New Keynesian hybrid

Phillips curve. Here we simulate data from the model

$$\begin{cases} \pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b^1 \pi_{t-1} + \gamma_b^2 \pi_{t-2} + \gamma_g g_t + \varepsilon_{\pi,t} \\ g_t = \beta_f E_t g_{t+1} + \beta_b^1 g_{t-1} + \beta_b^2 g_{t-2} - \beta_r (R_t - E_t \pi_{t+1}) + \varepsilon_{g,t} \\ R_t = (1-\rho)(\omega_\pi \pi_t + \omega_g g_t) + \rho R_{t-1} + \varepsilon_{R,t} \end{cases}$$

$$\begin{cases} \varepsilon_{\pi,t} = u_{\pi,t} \\ \varepsilon_{g,t} = \rho_g \varepsilon_{g,t-1} + u_{g,t} \\ \varepsilon_{R,t} = \rho_R \varepsilon_{R,t-1} + u_{R,t} \end{cases} \qquad \mathbf{u}_t \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} 0.5^2 & 0 & 0 \\ 0 & 0.288^2 & 0 \\ 0 & 0 & 0.252^2 \end{bmatrix} \right)$$

for different combinations of parameters to be made explicit shortly. The Phillips and IS curves are modified to include an extra lag than what is conventional to generate small distortions to the canonical specification and check the robustness of PMD and GMM to dynamic misspecification. Hence, some of the simulations are conducted with the standard specification where $\gamma_f + \gamma_b^1 = 1$ and $\gamma_b^2 = 0$ (and similarly for the IS curve parameters). In other experiments, $\gamma_f + \gamma_b^1 + \gamma_b^2 = 1$ and $\gamma_b^1 = \gamma_b^2$ (and similarly for the IS curve parameters) to induce additional serial correlation.

Most of the parameter choices are borrowed from Lindé (2005) and we refer the reader there for a more careful justification of these choices. We investigate three primary different combinations of parameters:

1. $\gamma_f = \beta_f = 0.7; \gamma_b^1 = \beta_b^1 = 0.3$ or $\gamma_b^1 = \gamma_b^2 = \beta_b^1 = \beta_b^2 = 0.15; \gamma_g = 0.13$; and $\beta_r = 0.09$
2. $\gamma_f = \beta_f = 0.5; \gamma_b^1 = \beta_b^1 = 0.5$ or $\gamma_b^1 = \gamma_b^2 = \beta_b^1 = \beta_b^2 = 0.25; \gamma_g = 0.25$; and $\beta_r = 0.30$
3. $\gamma_f = \beta_f = 0.3; \gamma_b^1 = \beta_b^1 = 0.7$ or $\gamma_b^1 = \gamma_b^2 = \beta_b^1 = \beta_b^2 = 0.35; \gamma_g = 0.40$; and $\beta_r = 1$

The Taylor rule parameters are the same in all cases with $\rho = 0.5; \omega_\pi = 1.5$ and $\omega_g = 0.5$ and the shock processes are allowed to take the two pairs of values $\rho_g = 0.5$ and $\rho_R = 0.8$ or $\rho_g = \rho_R = 0$. The latter case is included as a benchmark since for this specification estimation by GMM using as instruments lagged values of the endogenous variables (including lags of R_t) is correct and should provide estimates of the parameters close to the theoretical values. Like Lindé (2005), we experimented by allowing R_t to be part of the instrument set originally. However, distortions to the GMM estimates were so considerable with respect to the PMD estimates that we decided to include estimates that only use lagged values of R_t to present the performance of our estimator in the least favorable cases for it (although results that include R_t as an instrument are available upon request).

1,000 Monte Carlo runs are generated in all 36 different cases, with results summarized in tables 2.1-2.3. Each run is initialized with 500 burn-in replications with which a sample of 200 observations (as in Lindé, 2005) is then generated. The DGP corresponding to this model is of order two and thus we experimented with $H = 2$ and H selected with Hall et al.'s (2007) information criterion. GMM estimates based on the same lags are computed for comparison purposes. The lag length of the first stage local projections is automatically selected by AIC_C.

It would be very tedious to comment on each of the numerous cases investigated but some general lessons are apparent. First, when the shocks are *i.i.d.* (so that there are no distortions to the internal dynamics of the model) and we examine the case we label "Benchmark" (with the traditional dynamic specification), both PMD and GMM provide good estimates although estimates of the output gap parameter of the Phillips curve tend to be somewhat downward biased with GMM but to a much lesser extent with PMD. Virtually in all cases, PMD estimates are more efficient than their GMM counterparts.

Whether the dynamic structure is modified by allowing serial correlation in the shocks, richer dynamics in the Phillips curve, or richer dynamics in the IS curve, both PMD and GMM have more difficulty in obtaining accurate estimates of the parameters, specially the output gap parameter. For example, in table 2.1 the additional serial correlation in the structural inflation Euler equation is enough to cause estimates of the output parameter to flip sign, although more generally, we simply observed estimates that were downward biased. Distortions to the degree of forward/backward looking behavior of the Phillips curve were, on the other hand, much more muted although the distortions obtained with GMM tend to be considerable larger than with PMD. In these cases, the parameter estimates changed quite a bit with the number of instruments included.

Overall, while PMD was not a universal panacea for every foreseeable type of misspecification, we obtained estimates that had a smaller bias than GMM in the majority of the cases. When the model was correctly specified, there was little difference between the methods but even here PMD was less biased and provided more efficient estimates. The introduction of relatively small modifications in the dynamic behavior of the model was enough to generate considerable distortions in the estimation of the output gap parameter, which plays a very prominent role in this literature. Almost in every case considered, the distortion caused the parameter to be downward biased. PMD mitigates this bias somewhat with respect to GMM but not to the extent that would have been desirable.

5 Empirical Application: Fuhrer and Olivei (2005) Revisited

Estimating the Phillips and IS curves by limited-information methods is difficult due to the poor small-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 to improve small sample performance. They dub this procedure "optimal instruments" GMM ($OI-$GMM) and explore its properties relative to conventional GMM and MLE estimators with Monte Carlo experiments.

We apply PMD to the same examples Fuhrer and Olivei (2005) analyze to provide the reader a context of comparison for our method. The basic specification is (using the same notation as in Fuhrer and Olivei, 2005):

$$z_t = (1-\mu)\, z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t \tag{22}$$

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 (22), 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 article.

Table 3 summarizes the empirical estimates of the output Euler equation (top panel) and the inflation Euler equation (bottom panel), which correspond to tables 4 and 5 respectively in Fuhrer and Olivei (2005). For each Euler equation, we report the original GMM, MLE, and $OI-$GMM estimates and below these, we include the PMD results based on choosing H with Hall et al.'s (2007) information criterion. Figure 1 displays the estimates of μ (top row) and γ (bottom row) in expression (22) as a function of h with two-standard error bands.

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 the top panel in table 3. PMD estimates of μ are close to GMM estimates but with similar standard errors, and not very different from MLE or OI-GMM. On the other hand, PMD estimates for γ are slightly larger in magnitude, of the correct sign and statistically significant. However, as figure 1 shows, while the estimates of μ appear to be somewhat stable to the choice of H, the estimates of γ are positive for any $H < 7$. This suggests that estimates of γ should be taken with caution as the model is likely dynamically misspecified.

Estimates of the inflation Euler equation, reported in the bottom panel of table 3, follow a similar pattern. For all three specifications, μ and γ are estimated to be similar to the GMM estimates but in all three specifications, the misspecification tests (not reported for brevity) reject the model. The last three columns of figure 1 show that while estimates of μ are relatively stable, estimates of γ for the HP and ST specifications are virtually negative for any H. The RULC specification suggests γ is mostly positive (with γ negative only for $H = 3$ and 4). Overall, the results suggests caution since every indication (from the overidentifying restrictions tests to the plots of the parameter estimates as a function of H) is that the model is dynamically misspecified.

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

6 Conclusion

This paper introduces a convenient and numerically robust estimator for models whose likelihood score functions are nonlinear in the parameters (and hence require numerical or simulation routines) but with a mapping between Wold coefficients and parameters that is linear and hence can be exploited with a two-step minimum distance procedure. Linear or linearized Euler equations common in DSGE models and VARMA models are two examples of models fitting this description so that macroeconomic applications are the most natural environment where our estimator can be advantageous.

The paper makes several contributions. Our estimator is in the same family as Sbordone's (2002) and Rotemberg and Woodford's (1997), however, we are careful to derive its asymptotic properties formally. In doing so, we provide asymptotic results for Jordà's (2005) local projection estimator in fairly general environments. The small sample properties of our estimator are very encouraging in terms of consistency and efficiency with respect to maximum likelihood and GMM.

Throughout the paper we have made a number of simplifying assumptions that merit more analysis than space considerations permit in this paper. These include relaxing the assumption of homoscedasticity; investigating other models that share the same features as DSGE and VARMA models in the sense of having nonlinear likelihood scores but linear Wold mappings; examining nonlinear mappings (since the asymptotic theory presented here is cast in sufficiently general terms to accommodate such an extension); and a more careful asymptotic analysis where H is allowed to grow with the sample, which may prove useful in investigating problems of weak identification and limiting results about the efficiency of PMD relative to the Cramer-Rao lower bound. 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 simple, robust and with reasonable small sample properties.

7 Appendix

7.1 Definitions and Notation

We find it useful to begin by defining and collecting the notation that we use for the proofs of the propositions and lemmas introduced above. Specifically:

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

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

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

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

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

(vi) $\underset{kn\times kn}{\widehat{\Gamma}_k} = (T\text{-}k\text{-}H)^{-1}\sum_{t=k}^{T-H} X_{t,k}X'_{t,k}$

(vii) $\underset{kn\times n}{\widehat{\Gamma}_{1-k,h}} = (T-k-H)^{-1}\sum_{t=k}^{T-H} X_{t,k}\mathbf{y}'_{t+h}; h = 1, ..., H$

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

7.2 Proof of Proposition 1

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

$$\underset{n\times kn}{\widehat{A}}(k,h) = (\widehat{A}_1^h, ..., \widehat{A}_k^h) = \widehat{\Gamma}'_{1-k,h}\widehat{\Gamma}_k^{-1} \tag{23}$$

Notice that

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

where

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

Hence,

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

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

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

Hence

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

where

$$\begin{aligned} U_{1T} &= \left\{(T-k-h^{-1}) \sum_{t=k}^{T-h} \left(\sum_{j=k+1}^{\infty} A_j^h \mathbf{y}_{t-j} \right) X'_{t,k} \right\} \qquad (24) \\ U_{2T} &= \left\{(T-k-h^{-1}) \sum_{t=k}^{T-h} \mathbf{u}_{t+h} X'_{t,k} \right\} \\ U_{3T} &= \left\{(T-k-h^{-1}) \sum_{t=k}^{T-h} \left(\sum_{j=1}^{h} B_j \mathbf{u}_{t+h-j} \right) X'_{t,k} \right\} \end{aligned}$$

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* $\mathbf{u}_{t+h}$ *and* $X'_{t,k}$ *are independent, so that their covariance is zero. Formally and following similar derivations in Lewis and Reinsel (1985)*

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

by independence. Hence

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

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

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

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

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

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

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

Hence

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

by recursive substitution. Thus

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

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

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

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

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

instead.

7.3 Proof of Proposition 2

Notice that

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

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

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

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

Hence, the strategy of the proof will consist in showing that the first term in the sum above vanishes in probability so that,

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

and then all we need to do is show that this last term is asymptotically normal. First we prove the convergence in probability result in this last expression. Define, U_{1T} as in expression (24),

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

then

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

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

which implies

$$(T-k-h)^{1/2}\,vec\left[\widehat{A}(k,h)-A(k,h)\right]-(T-k-h)^{1/2}\,vec\left[U_{2T}^{*}\Gamma_k^{-1}\right]=W_{1T}+W_{2T}+W_{3T}$$

where,

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

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

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

0; $k, T \to \infty$ from assumption (iii) and using similar derivations as in the proof of consistency with s being a generic constant. Hence $W_{1T} \xrightarrow{p} 0$.

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

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

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

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

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

and substituting this result into their proof.

The asymptotic normality result then follows directly from Lewis and Reinsel (1985) by redefining

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

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

7.4 Proof of Lemma 3

Since $\widehat{\mathbf{b}}_T \xrightarrow{p} \mathbf{b}_0$, then

$$f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right) \xrightarrow{p} f\left(\mathbf{b}_0; \boldsymbol{\gamma}\right)$$

by the continuous mapping theorem since by assumption (iv), $f(.)$ is continuous. Furthermore and given assumption (i)

$$\widehat{Q}_T(\boldsymbol{\gamma}) = f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right)' \widehat{W} f\left(\widehat{\mathbf{b}}_T; \boldsymbol{\gamma}\right) \xrightarrow{p} f\left(\mathbf{b}_0; \boldsymbol{\gamma}\right)' \widehat{W} f\left(\mathbf{b}_0; \boldsymbol{\gamma}\right) \equiv Q_0(\boldsymbol{\gamma})$$

which is a quadratic expression that is maximized at $\boldsymbol{\gamma}_0$*. 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,* $Q_0(\boldsymbol{\gamma})$ *is obviously a continuous function. The last thing to show is that*

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

uniformly.

For compact Θ *and continuous* $Q_0(\boldsymbol{\gamma})$*, Lemma 2.8 in Newey and McFadden (1994) provides that this condition holds if and only if* $\widehat{Q}_T(\boldsymbol{\gamma}) \xrightarrow{p} Q_0(\boldsymbol{\gamma})$ *for all* $\boldsymbol{\gamma}$ *in* Θ *and* $\widehat{Q}_T(\boldsymbol{\gamma})$ *is stochastically equicontinuous. The former has already been established, so it remains to show stochastic equicontinuity of* $\widehat{Q}_T(\boldsymbol{\gamma})$*.*[2] *However* $\widehat{Q}_T(\boldsymbol{\gamma})$ *in this paper the function* $f(.)$ *is rather trivial (linear in the parameters) so that uniform convergence itself is rather straight-forward to establish. In general, we directly assume here that stochastic continuity holds and we refer the reader to Andrews (1994, 1995) for examples and sets of specific conditions for more general forms of* $f(.)$*.*

7.5 Proof of Lemma 4

Under assumption (iii) $\mathbf{b}_0$ *and* $\boldsymbol{\gamma}_0$ *are in the interior of their parameter spaces and by assumption (ii)* $\widehat{\mathbf{b}}_T \xrightarrow{p} \mathbf{b}_0$, $\widehat{\boldsymbol{\gamma}}_T \xrightarrow{p} \boldsymbol{\gamma}_0$*. Further, by assumption (iv),* $f(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma})$ *is continuously differentiable in a neighborhood of* $\mathbf{b}_0$ *and* $\boldsymbol{\gamma}_0$ *and hence* $\widehat{\boldsymbol{\gamma}}_T$ *solves the first order conditions of the minimum-distance problem*

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

which are

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

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

By assumption (iv), these first order conditions can be expanded about $\boldsymbol{\gamma}_0$ in mean value expansion

$$f(\widehat{\mathbf{b}}_T;\widehat{\boldsymbol{\gamma}}_T) = f(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma}_0) + F_\gamma\left(\widehat{\mathbf{b}}_T;\overline{\boldsymbol{\gamma}}\right)(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0)$$

where $\overline{\boldsymbol{\gamma}} \in [\widehat{\boldsymbol{\gamma}}_T, \boldsymbol{\gamma}_0]$. Similarly, a mean value expansion of $f(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma}_0)$ around $\mathbf{b}_0$ is

$$f(\widehat{\mathbf{b}}_T;\boldsymbol{\gamma}_0) = \mathbf{f}(\mathbf{b}_0;\boldsymbol{\gamma}_0) + F_{\mathbf{b}}\left(\overline{\mathbf{b}};\boldsymbol{\gamma}_0\right)\left(\widehat{\mathbf{b}}_T - \mathbf{b}_0\right)$$

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

$$\begin{aligned}\sqrt{T}f(\widehat{\mathbf{b}}_T;\widehat{\boldsymbol{\gamma}}_T) &= \sqrt{T}f(\mathbf{b}_0;\boldsymbol{\gamma}_0) + F_\gamma\left(\widehat{\mathbf{b}}_T;\overline{\boldsymbol{\gamma}}\right)\sqrt{T}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) + \\ &\quad F_{\mathbf{b}}\left(\overline{\mathbf{b}};\boldsymbol{\gamma}_0\right)\sqrt{T}\left(\widehat{\mathbf{b}}_T - \mathbf{b}_0\right)\end{aligned}$$

Since $\overline{\mathbf{b}} \in [\widehat{\mathbf{b}}_T, \mathbf{b}_0]$, $\overline{\boldsymbol{\gamma}} \in [\widehat{\boldsymbol{\gamma}}_T, \boldsymbol{\gamma}_0]$ and $\widehat{\mathbf{b}}_T \xrightarrow{p} \mathbf{b}_0$, $\widehat{\boldsymbol{\gamma}}_T \xrightarrow{p} \boldsymbol{\gamma}_0$ then, along with assumption (iv), we have

$$\begin{aligned}F_\gamma\left(\widehat{\mathbf{b}}_T;\overline{\boldsymbol{\gamma}}\right) \xrightarrow{p} F_\gamma\left(\mathbf{b}_0;\boldsymbol{\gamma}_0\right) &= F_\gamma \\ F_{\mathbf{b}}\left(\overline{\mathbf{b}};\boldsymbol{\gamma}_0\right) \xrightarrow{p} F_{\mathbf{b}}(\mathbf{b}_0;\boldsymbol{\gamma}_0) &= F_{\mathbf{b}}\end{aligned}$$

and hence

$$\sqrt{T}f(\widehat{\mathbf{b}}_T;\widehat{\boldsymbol{\gamma}}_T) = \sqrt{T}f(\mathbf{b}_0;\boldsymbol{\gamma}_0) + F_\gamma\sqrt{T}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) + F_{\mathbf{b}}\sqrt{T}\left(\widehat{\mathbf{b}}_T - \mathbf{b}_0\right) + o_p(1)$$

In addition, by assumption (i) $\widehat{W} \xrightarrow{p} W$ and notice that $f(\mathbf{b}_0, \boldsymbol{\gamma}_0) = 0$, which combined with the first order conditions and the mean value expansions described above, allow us to write

$$-F_\gamma' W\left[F_\gamma\sqrt{T}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) + F_{\mathbf{b}}\sqrt{T}\left(\widehat{\mathbf{b}}_T - \mathbf{b}_0\right)\right] = o_p(1)$$

Since we know that

$$\sqrt{T}\left(\widehat{\mathbf{b}}_T - \mathbf{b}_0\right) \xrightarrow{d} N(0, \Omega_b)$$

by proposition 2, then

$$\sqrt{T}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) \xrightarrow{d} -\left(F_\gamma' W F_\gamma\right)^{-1}\left(F_\gamma' W F_{\mathbf{b}}\right)\sqrt{T}\left(\widehat{\mathbf{b}}_T - \mathbf{b}_0\right)$$

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

$$\sqrt{T}\left(\widehat{\gamma}_T - \gamma_0\right) \xrightarrow{d} N\left(0, \Omega_\gamma\right)$$
$$\Omega_\gamma = \left(F_\gamma' W F_\gamma\right)^{-1}\left(F_\gamma' W F_{\mathbf{b}} \Omega_{\mathbf{b}} F_{\mathbf{b}}' W F_\gamma\right)\left(F_\gamma' W F_\gamma\right)^{-1}$$

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

$$\Omega_\gamma = \left(F_\gamma' W F_\gamma\right)^{-1}$$
$$W = \left(F_{\mathbf{b}} \Omega_{\mathbf{b}} F_{\mathbf{b}}'\right)^{-1}$$

References

An, Sungbae and Frank Schorfheide (2007) "Bayesian Analysis of DSGE Models," *Econometric Reviews,* 26(2-4): 113-172.

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.

Andrews, Donald W. K. and James H. Stock (2007) "Testing with Many Weak Instruments," *Journal of Econometrics*, 138(1): 24-46.

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

Canova, Fabio (2007) **Methods for Applied Macroeconomic Research.** Princeton: Princeton University Press.

Christiano, Lawrence J. and Wouter den Haan (1996) "Small-Sample Properties of GMM for Business Cycle Analysis," *Journal of Business and Economic Statistics,* 14(3): 309-327.

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.

Cogley, Timothy and Argia M. Sbordone (2008) "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review,* 98(5): 2101-2126.

DeJong, David N., Beth F. Ingram and Charles H. Whiteman (2000) "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics,* 98(2): 203-223.

Demiralp, Selva and Kevin D. Hoover (2003) "Searching for the Causal Structure of a Vector Autoregression," *Oxford Bulletin of Economics and Statistics,* 65(Supplement): 745-767.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent and Mark W. Watson (2007) "A, B, Cs (and Ds) of Understanding VARs," *American Economic Review,*97(3): 1021-1026.

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

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

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

Granger, Clive W. J. and Norman R. Swanson (1997) "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association,* 92(437): 357-367.

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

Harville, David A. (1997) **Matrix Algebra From A Statisticians Perspective.** Springer: New York.

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

Jondeau, Eric and Hervé Le Bihan (2007) "Examining Bias in Estimators of Linear Rational Expectations Models Under Misspecification," *Journal of Econometrics,* 143(2): 375-395.

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

Jordà, Òscar (2009) "Simultaneous Confidence Regions for Impulse Responses," *Review of Economics and Statistics, 91*(3): forthcoming.

Kuersteiner, Guido M. (2005) "Automatic Inference for Infinite Order Vector Autoregressions," *Econometric Theory,* 21: 85-115.

Kurmann, André (2005) "Quantifying the Uncertainty about a Forward-Looking New Keynesian Pricing Model," *Journal of Monetary Economics,* 52(6): 1119-1134.

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.

Levin, Andrew T. and John C. Williams (2003) "Robust Monetary Policy with Competing Reference Models," *Journal of Monetary Economics,* 50(5): 945-975.

Lindé, Jesper (2005) "Estimating New Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach," *Journal of Monetary Economics,* 52(6): 1135-1149.

Mavroeidis, Sophocles (2004) "Weak Identification of Forward-Looking Models in Monetary Economics," *Oxford Bulletin of Economics and Statistics,* 66(S1): 609-635.

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.

Rudd, Jeremy and Karl Whelan (2005) "New Test of the New-Keynesian Phillips Curve," *Journal of Monetary Economics,* 52(6): 1167-1181.

Rudd, Jeremy and Karl Whelan (2006) "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?" *American Economic Review,* 96(1): 303-320.

Sargent, Thomas J. (1989) "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy,* 97(2): 251-287.

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

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.

Wooldridge, Jeffrey M. (1994) "Estimation and Inference for Dependent Processes," in **Handbook of Econometrics, v. 4**, Robert F. Engle and Daniel L. McFadden, (eds.). Amsterdam: North Holland.

Table 1 – ARMA (1,1) Monte Carlo Experiments

	Case 1: $(\rho, \theta) = (0.25, 0.5)$				Case 2: $(\rho, \theta) = (0.5, 0.25)$				Case 3: $(\rho, \theta) = (0, 0.5)$				Case 4: $(\rho, \theta) = (0.5, 0)$			
	h = 2		h = 5		h = 2		h = 5		h = 2		h = 5		h = 2		h = 5	
T = 50	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ
PMD	0.23	0.49	0.25	0.44	0.46	0.23	0.47	0.17	-0.06	0.56	0.06	0.40	0.47	0.04	0.43	0.03
	(0.22)	(0.20)	(0.20)	(0.19)	(0.19)	(0.20)	(0.18)	(0.19)	(0.36)	(0.32)	(0.27)	(0.25)	(0.28)	(0.30)	(0.24)	(0.26)
MLE	0.22	0.52	0.23	0.52	0.45	0.29	0.44	0.27	-	-	-	-	-	-	-	-
	(0.21)	(0.18)	(0.20)	(0.18)	(0.20)	(0.20)	(0.20)	(0.21)	-	-	-	-	-	-	-	-
T = 100	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ
PMD	0.24	0.50	0.25	0.47	0.48	0.23	0.47	0.23	-0.03	0.54	0.04	0.45	0.49	0.01	0.45	0.03
	(0.15)	(0.14)	(0.15)	(0.13)	(0.13)	(0.14)	(0.13)	(0.14)	(0.24)	(0.21)	(0.19)	(0.18)	(0.19)	(0.20)	(0.17)	(0.18)
MLE	0.25	0.51	0.24	0.51	0.48	0.27	0.47	0.25	-	-	-	-	0.49	-0.02	0.47	0.03
	(0.14)	(0.13)	(0.14)	(0.13)	(0.14)	(0.14)	(0.14)	(0.15)	-	-	-	-	(0.17)	(0.20)	(0.18)	(0.20)
T = 400	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ	ρ	θ
PMD	0.25	0.51	0.25	0.50	0.50	0.25	0.49	0.26	-0.01	0.51	0.00	0.50	0.50	0.01	0.50	0.00
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)	(0.11)	(0.10)	(0.10)	(0.09)	(0.09)	(0.10)	(0.08)	(0.09)
MLE	0.25	0.50	0.25	0.50	0.50	0.25	0.49	0.26	0.04	0.50	0.00	0.50	0.49	0.01	0.49	0.01
	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)

Notes: 1,000 Monte Carlo replications. First stage regression lag-length chosen automatically by AIC_C . Standard errors in parenthesis calculated with PMD/MLE analytic formulas. 500 burn-in observations used to initialize each run. Blank entries are due to lack of convergence of the maximum likelihood estimator in some of the runs and are meant to indicate numerical difficulties for the particular parameter combinations where they occur.

Table 2.1 Monte Carlo Comparison: GMM vs. PMD.

Case 1: $w_f = \beta_f = 0.7; w_b = \beta_b = 0.3; \gamma = 0.13; \beta_r = 0.09; \rho = 0.5; \gamma_\pi = 1.50; \gamma_y = 0.5$

D.G.P.

$$\begin{cases} \pi_t = w_f E_t \pi_{t+1} + w_b^1 \pi_{t-1} + w_b^2 \pi_{t-2} + \gamma y_t + \epsilon_{\pi,t} \\ y_t = \beta_f E_t y_{t+1} + \beta_b^1 y_{t-1} + \beta_b^2 y_{t-2} - \beta_r (R_t - E_t \pi_{t+1}) + \epsilon_{y,t} \\ R_t = (1-\rho)(\gamma_\pi \pi_t + \gamma_y y_t) + \rho R_{t-1} + \epsilon_{R,t} \end{cases} \qquad \begin{cases} \epsilon_{\pi,t} = u_{\pi,t} & u_{\pi,t} \backsim N(0, 0.5^2) \\ \epsilon_{y,t} = \rho_y \epsilon_{y,t-1} + u_{y,t} & u_{y,t} \backsim N(0, 0.288^2) \\ \epsilon_{R,t} = \rho_R \epsilon_{R,t-1} + u_{R,t} & u_{R,t} \backsim N(0, 0.252^2) \end{cases}$$

	$\rho_g = \rho_R = 0$				$\rho_g = 0.5; \rho_R = 0.8$			
	PMD		**GMM**		**PMD**		**GMM**	
Benchmark	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.7$	0.703	0.689	0.697	0.678	0.718	0.692	0.705	0.689
	(0.038)	(0.030)	(0.041)	(0.036)	(0.026)	(0.021)	(0.032)	(0.030)
$w_b^1 = 0.3; w_b^2 = 0$	0.297	0.312	0.303	0.322	0.281	0.307	0.295	0.311
	(0.038)	(0.071)	(0.041)	(0.036)	(0.026)	(0.021)	(0.032)	(0.030)
$\gamma = 0.13$	0.127	0.102	0.108	0.098	0.078	0.074	0.078	0.064
	(0.097)	(0.030)	(0.119)	(0.102)	(0.036)	(0.027)	(0.048)	(0.045)
	PMD		**GMM**		**PMD**		**GMM**	
Lagged PC	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.7$	0.332	0.503	0.507	0.267	0.422	0.453	0.698	0.369
	(0.411)	(0.135)	(0.350)	(0.219)	(0.313)	(0.109)	(0.295)	(0.196)
$w_b^1 = w_b^2 = 0.15$	0.140	0.160	0.080	0.098	0.154	0.192	0.071	0.116
	(0.100)	(0.051)	(0.082)	(0.075)	(0.079)	(0.040)	(0.072)	(0.064)
$\gamma = 0.13$	-0.063	-0.071	-0.039	-0.068	0.107	0.079	0.082	0.118
	(0.216)	(0.100)	(0.184)	(0.174)	(0.087)	(0.042)	(0.083)	(0.078)
	PMD		**GMM**		**PMD**		**GMM**	
Lagged IS	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.7$	0.708	0.682	0.697	0.680	0.714	0.684	0.704	0.682
	(0.041)	(0.030)	(0.043)	(0.036)	(0.026)	(0.019)	(0.031)	(0.029)
$w_b^1 = 0.3; w_b^2 = 0$	0.291	0.317	0.303	0.320	0.285	0.316	0.296	0.317
	(0.041)	(0.030)	(0.043)	(0.036)	(0.026)	(0.019)	(0.031)	(0.029)
$\gamma = 0.13$	0.082	0.082	0.067	0.043	0.043	0.043	0.040	0.037
	(0.151)	(0.101)	(0.189)	(0.151)	(0.052)	(0.036)	(0.065)	(0.062)

Notes: 1,000 Monte Carlo replications. Each run initialized with 500 burn-in replications later disregarded. Sample size T = 200. Monte Carlo median values of the parameter estimates and the associated standard errors reported. “Lagged PC” refers to when the DGP consists of a Phillips Curve with first and second lag inflation terms. Similarly, “Lagged IS” refers to when the DGP consists of an IS curve with first and second lag output gap terms. *h = 2* uses the first 2 horizons of the impulse response function when estimating with PMD (to obtain over-identification) and corresponds to using the first two lags of the variables as instruments when estimating by GMM. h^* refers to the optimal horizon selected by Hall et al.’s (2007) information criterion and varies with the model. “Benchmark” and “Lagged IS” cases impose $w_f + w_b = 1$. “Lagged PC” case estimates these parameters unconstrained (since in the DGP $w_f + w_b^1 + w_b^2 = 1$ but $w_f + w_b^1 \neq 1$).

Table 2.2 Monte Carlo Comparison: GMM vs. PMD.

Case 2: $w_f = \beta_f = 0.5; w_b = \beta_b = 0.5; \gamma = 0.25; \beta_r = 0.30; \rho = 0.5; \gamma_\pi = 1.50; \gamma_y = 0.5$

D.G.P.

$$\begin{cases} \pi_t = w_f E_t \pi_{t+1} + w_b^1 \pi_{t-1} + w_b^2 \pi_{t-2} + \gamma y_t + \epsilon_{\pi,t} \\ y_t = \beta_f E_t y_{t+1} + \beta_b^1 y_{t-1} + \beta_b^2 y_{t-2} - \beta_r (R_t - E_t \pi_{t+1}) + \epsilon_{y,t} \\ R_t = (1-\rho)(\gamma_\pi \pi_t + \gamma_y y_t) + \rho R_{t-1} + \epsilon_{R,t} \end{cases} \quad \begin{cases} \epsilon_{\pi,t} = u_{\pi,t} & u_{\pi,t} \backsim N(0, 0.5^2) \\ \epsilon_{y,t} = \rho_y \epsilon_{y,t-1} + u_{y,t} & u_{y,t} \backsim N(0, 0.288^2) \\ \epsilon_{R,t} = \rho_R \epsilon_{R,t-1} + u_{R,t} & u_{R,t} \backsim N(0, 0.252^2) \end{cases}$$

	$\rho_g = \rho_R = 0$				$\rho_g = 0.5; \rho_R = 0.8$			
	PMD		**GMM**		**PMD**		**GMM**	
Benchmark	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.5$	0.509	0.510	0.516	0.523	0.569	0.543	0.557	0.550
	(0.035)	(0.028)	(0.052)	(0.046)	(0.027)	(0.021)	(0.037)	(0.035)
$w_b^1 = 0.5; w_b^2 = 0$	0.491	0.490	0.484	0.476	0.431	0.456	0.443	0.450
	(0.035)	(0.028)	(0.052)	(0.046)	(0.027)	(0.021)	(0.037)	(0.035)
$\gamma = 0.25$	0.251	0.220	0.229	0.197	0.159	0.164	0.159	0.148
	(0.040)	(0.032)	(0.073)	(0.064)	(0.027)	(0.022)	(0.042)	(0.040)
	PMD		**GMM**		**PMD**		**GMM**	
Lagged PC	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.5$	0.356	0.509	0.619	0.318	0.556	0.560	0.888	0.581
	(0.385)	(0.124)	(0.328)	(0.205)	(0.166)	(0.079)	(0.207)	(0.152)
$w_b^1 = w_b^2 = 0.25$	0.228	0.251	0.084	0.131	0.280	0.345	0.085	0.187
	(0.113)	(0.052)	(0.098)	(0.078)	(0.071)	(0.036)	(0.078)	(0.065)
$\gamma = 0.25$	0.079	0.040	0.028	0.061	0.121	0.088	0.050	0.141
	(0.119)	(0.050)	(0.106)	(0.088)	(0.063)	(0.033)	(0.086)	(0.069)
	PMD		**GMM**		**PMD**		**GMM**	
Lagged IS	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.5$	0.547	0.535	0.534	0.542	0.571	0.557	0.559	0.554
	(0.057)	(0.032)	(0.054)	(0.045)	(0.030)	(0.019)	(0.036)	(0.033)
$w_b^1 = 0.5; w_b^2 = 0$	0.452	0.465	0.466	0.457	0.429	0.442	0.441	0.446
	(0.057)	(0.032)	(0.054)	(0.045)	(0.030)	(0.019)	(0.036)	(0.033)
$\gamma = 0.25$	0.126	0.145	0.142	0.118	0.093	0.109	0.098	0.090
	(0.103)	(0.048)	(0.093)	(0.078)	(0.041)	(0.025)	(0.049)	(0.046)

Notes: 1,000 Monte Carlo replications. Each run initialized with 500 burn-in replications later disregarded. Sample size T = 200. Monte Carlo median values of the parameter estimates and the associated standard errors reported. "Lagged PC" refers to when the DGP consists of a Phillips Curve with first and second lag inflation terms. Similarly, "Lagged IS" refers to when the DGP consists of an IS curve with first and second lag output gap terms. *h = 2* uses the first 2 horizons of the impulse response function when estimating with PMD (to obtain over-identification) and corresponds to using the first two lags of the variables as instruments when estimating by GMM. h^* refers to the optimal horizon selected by Hall et al.'s (2007) information criterion and varies with the model. "Benchmark" and "Lagged IS" cases impose $w_f + w_b = 1$. "Lagged PC" case estimates these parameters unconstrained (since in the DGP $w_f + w_b^1 + w_b^2 = 1$ but $w_f + w_b^1 \neq 1$).

Table 2.3 Monte Carlo Comparison: GMM vs. PMD.

Case 3: $w_f = \beta_f = 0.3; w_b = \beta_b = 0.7; \gamma = 0.40; \beta_r = 1; \rho = 0.5; \gamma_\pi = 1.50; \gamma_y = 0.5$

D.G.P.

$$\begin{cases} \pi_t = w_f E_t \pi_{t+1} + w_b^1 \pi_{t-1} + w_b^2 \pi_{t-2} + \gamma y_t + \epsilon_{\pi,t} \\ y_t = \beta_f E_t y_{t+1} + \beta_b^1 y_{t-1} + \beta_b^2 y_{t-2} - \beta_r (R_t - E_t \pi_{t+1}) + \epsilon_{y,t} \\ R_t = (1-\rho)(\gamma_\pi \pi_t + \gamma_y y_t) + \rho R_{t-1} + \epsilon_{R,t} \end{cases} \qquad \begin{cases} \epsilon_{\pi,t} = u_{\pi,t} & u_{\pi,t} \backsim N(0, 0.5^2) \\ \epsilon_{y,t} = \rho_y \epsilon_{y,t-1} + u_{y,t} & u_{y,t} \backsim N(0, 0.288^2) \\ \epsilon_{R,t} = \rho_R \epsilon_{R,t-1} + u_{R,t} & u_{R,t} \backsim N(0, 0.252^2) \end{cases}$$

	$\rho_g = \rho_R = 0$				$\rho_g = 0.5; \rho_R = 0.8$			
	PMD		**GMM**		**PMD**		**GMM**	
Benchmark	***h = 2***	***h* = 4***	***h = 2***	***h* = 4***	***h = 2***	***h* = 4***	***h = 2***	***h* = 4***
$w_f = 0.3$	0.300	0.303	0.327	0.340	0.418	0.419	0.410	0.410
	(0.054)	(0.029)	(0.076)	(0.071)	(0.042)	(0.031)	(0.056)	(0.055)
$w_b^1 = 0.7; w_b^2 = 0$	0.700	0.697	0.672	0.640	0.582	0.580	0.590	0.589
	(0.054)	(0.029)	(0.076)	(0.071)	(0.042)	(0.031)	(0.056)	(0.055)
$\gamma = 0.40$	0.402	0.400	0.372	0.349	0.310	0.304	0.288	0.283
	(0.051)	(0.027)	(0.079)	(0.074)	(0.031)	(0.022)	(0.054)	(0.053)
	PMD		**GMM**		**PMD**		**GMM**	
Lagged PC	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.3$	0.280	0.409	0.564	0.292	0.697	0.605	1.029	0.801
	(0.406)	(0.135)	(0.400)	(0.228)	(0.153)	(0.068)	(0.174)	(0.127)
$w_b^1 = w_b^2 = 0.35$	0.289	0.310	0.133	0.160	0.259	0.419	0.058	0.172
	(0.112)	(0.053)	(0.107)	(0.083)	(0.093)	(0.039)	(0.092)	(0.071)
$\gamma = 0.40$	0.183	0.129	0.130	0.139	0.102	0.089	0.007	0.078
	(0.102)	(0.042)	(0.101)	(0.077)	(0.062)	(0.029)	(0.082)	(0.062)
$\gamma = 0.40$	**PMD**		**GMM**		**PMD**		**GMM**	
Lagged IS	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***	***h = 2***	***h* = 6***
$w_f = 0.3$	0.378	0.403	0.405	0.440	0.468	0.436	0.446	0.461
	(0.089)	(0.038)	(0.078)	(0.060)	(0.044)	(0.026)	(0.052)	(0.045)
$w_b^1 = 0.7; w_b^2 = 0$	0.621	0.596	0.595	0.559	0.532	0.561	0.554	0.539
	(0.089)	(0.038)	(0.078)	(0.060)	(0.044)	(0.026)	(0.052)	(0.045)
$\gamma = 0.40$	0.280	0.248	0.256	0.202	0.196	0.204	0.201	0.180
	(0.103)	(0.041)	(0.092)	(0.071)	(0.041)	(0.028)	(0.056)	(0.050)

Notes: 1,000 Monte Carlo replications. Each run initialized with 500 burn-in replications later disregarded. Sample size T = 200. Monte Carlo median values of the parameter estimates and the associated standard errors reported. "Lagged PC" refers to when the DGP consists of a Phillips Curve with first and second lag inflation terms. Similarly, "Lagged IS" refers to when the DGP consists of an IS curve with first and second lag output gap terms. *h = 2* uses the first 2 horizons of the impulse response function when estimating with PMD (to obtain over-identification) and corresponds to using the first two lags of the variables as instruments when estimating by GMM. h^* refers to the optimal horizon selected by Hall et al.'s (2007) information criterion and varies with the model. "Benchmark" and "Lagged IS" cases impose $w_f + w_b = 1$. "Lagged PC" case estimates these parameters unconstrained (since in the DGP $w_f + w_b^1 + w_b^2 = 1$ but $w_f + w_b^1 \neq 1$).

Table 3 – GMM, MLE, Optimal Instruments GMM and PMD: A Comparison

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

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

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^* = 11)	**HP**	0.50 (0.027)	-0.021 (0.011)
PMD (h^* = 11)	**ST**	0.50 (0.027)	-0.020 (0.011)

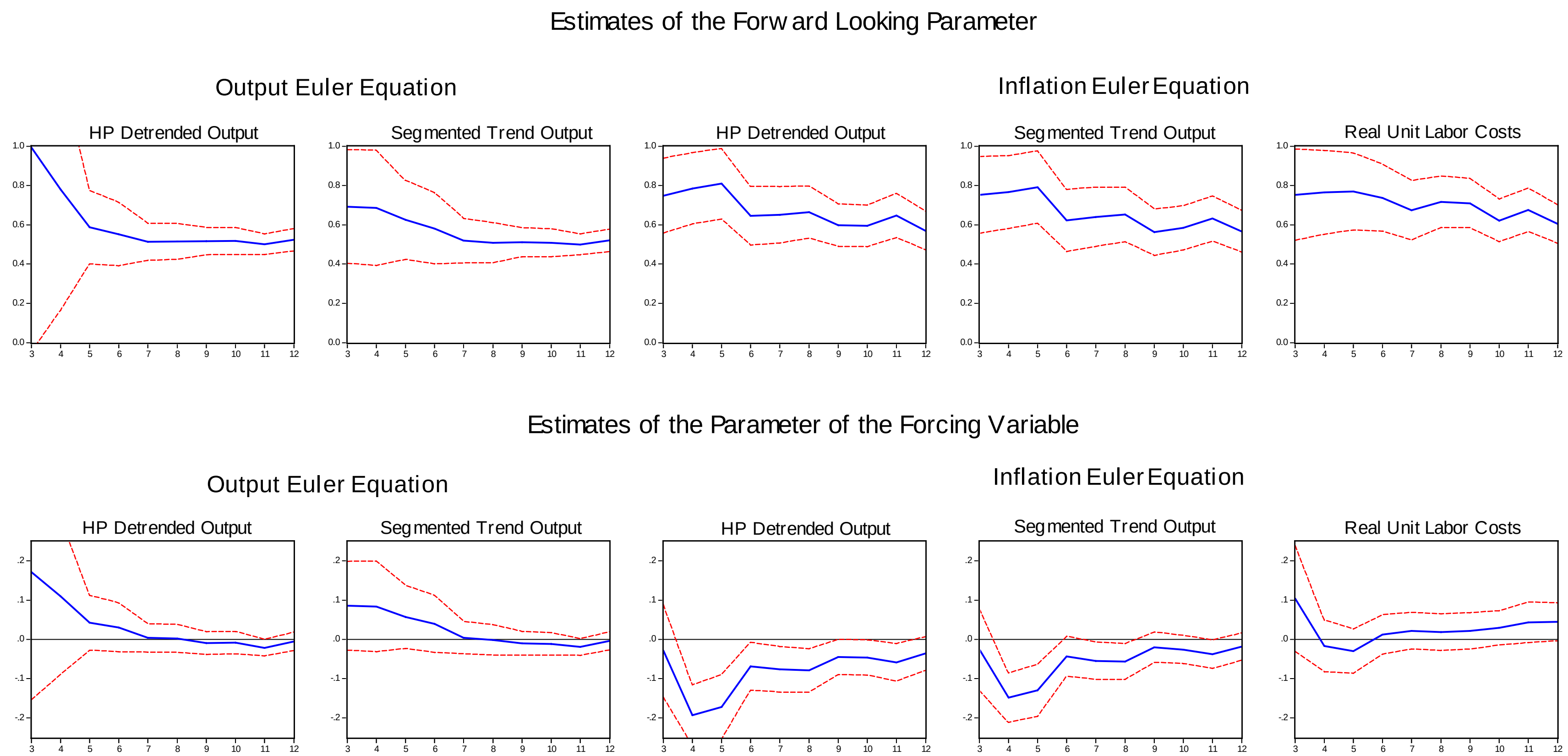
Notes: z_t is a measure of the output gap, x_t is a measure of the real interest rate, and hence economic theory would predict $\gamma < 0$. GMM, MLE, and OI-GMM estimates correspond to estimates reported in Table 4 in Fuhrer and Olivei (2005). HP refers to Hodrick-Prescott filtered log of real GDP, and ST refers to log of real GDP detrended by a deterministic segmented trend. Optimal value of *h* for PMD determined by Hall et al.'s (2007) information criterion and displayed in parenthesis as h^*.

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

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^* = 9)	**HP**	0.60 (0.055)	-0.045 (0.022)
PMD (h^* = 10)	**ST**	0.58 (0.057)	-0.026 (0.018)
PMD (h^* = 10)	**RULC**	0.62 (0.056)	0.029 (0.022)

Notes: z_t is a measure of inflation, x_t is a measure of the output gap, and hence economic theory would predict $\gamma > 0$. GMM, MLE and OI-GMM estimates correspond to estimates reported in Table 5 in Fuhrer and Olivei (2005). HP refers to Hodrick-Prescott filtered log of real GDP, and ST refers to log of real GDP detrended by a deterministic segmented trend. RULC refers to real unit labor costs. Optimal value of *h* for PMD determined by Hall et al.'s (2007) information criterion and displayed in parenthesis as h^*.

Figure 1 – PMD Estimates as a function of the impulse response horizon for Output and Inflation Euler Equations: 1966Q1 – 2001Q4



Notes: first two columns refer to the parameter estimates of the output Euler equation, last three columns to the inflation Euler equation. First row refers to the estimates of the μ parameter, where as second row refers to estimates of the γ parameter in expression (20) in the text. Two standard error bands around estimates reported.