This version, March 2006

Projection Minimum Distance: An Estimator for Dynamic Macroeconomic Models*

Abstract

This paper introduces a simple, two-step estimator for rational expectations models and time series models in general. We call this estimator projection minimum distance (PMD) and show that it is consistent and asymptotically normal. PMD provides consistent estimates of parameters even when the economic model is misspecified. In addition, it provides a simple chi-square specification test and a statistically based metric of closeness between the impulse responses implied by the model and those from the data. We show how PMD relates to GMM and how invalid/weak instrument problems in GMM are easily resolved by PMD. PMD is simple to implement as it only involves two simple least squares steps and can be generalized to more complex and nonlinear environments.

- *Keywords:* impulse response, local projection, rational expectations, minimum chi-squared estimation.
- *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 Timothy Cogley, James Hamilton, Peter Hansen, Kevin Hoover, Frank Wolak and seminar participants at Duke University, the Federal Reserve Bank of New York, Federal Reserve Bank of San Francisco, Federal Reserve Bank of St. Louis, Stanford University, the University of California, Davis, University of Kansas and the Winter Meetings of the American Economic Association, 2006 in Boston for useful comments and suggestions. Jordà is thankful for the hospitality of the Federal Reserve Bank of San Francisco during the creation of this paper.

1 Introduction

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

This paper introduces new methods to estimate and evaluate structural, dynamic, stochastic macroeconomic models. The method, which we label projection minimum distance (PMD), is a two-step estimator. In the first step, we compute the impulse responses from the data semiparametrically with the local projections estimator introduced by Jordà (2005). Next, we represent the stable solution of the model in terms of its Wold representation and obtain the mapping between the structural parameters and the Wold coefficients by the method of undetermined coefficients (see Christiano, 2002). Hence, in the second step, the structural parameters of the model are effectively estimated by minimizing the distance between the impulse responses from the data and those implied by the model. The resulting estimator is based on a minimum chi-square estimator (Ferguson, 1958) and belongs to the broader family of minimum distance estimators of which GMM is also a member of the class.

However, PMD has important advantages that distinguish it from GMM and other commonly used estimators. Specifically, PMD consists of two, simple, least-squares steps, and therefore, is easily implementable. The principle behind PMD consists in minimizing the distance between the data's and the model's impulse responses – the dimension of fit macroeconomic researchers most care about. Two implications follow from this principle: First, we provide an overall misspecification test based on overidentifying restrictions that is distributed chi-square. Effectively, this test is the formal equivalent of the "eyeball" determination of when a model matches the relevant dynamics of the data. Secondly, PMD is designed to provide consistent estimates even when the model is dynamically misspecified. We show that the common practice of using lags of the endogenous variables as instruments in GMM can only be justified by the internal dynamics of the data, not the model. Otherwise, GMM presents and invalid instrument problem that results in inconsistent estimates. Fuhrer and Olivei (2004) explain this problem in more detail.

Econometrically, the paper has three contributions. First, it is critical for the consistency properties of PMD to obtain consistent, first-step estimates of the impulse responses from the data. This motivates our preference for local projections over approximations based on vector autoregressions (VARs). Consequently, we begin by deriving consistency and asymptotic normality results for the local projection estimates. Second and unlike classical minimum distance, the minimum chi-square step is based on a function of the structural parameters that depends on sample estimates. Consequently, we derive the consistency and asymptotic normality of the minimum chi-square step and appropriately incorporate into the asymptotic variance-covariance matrix the uncertainty generated in the first-step. Hence, we show that a misspecification test of overidentifying restrictions is distributed chi-square. Finally, we show that PMD is consistent even when the economic model is dynamically misspecified and how our model corrects the invalid instrument problem that GMM suffers.

We introduce PMD and the main results in the context of a flexible, linear state-space representation of a dynamic rational expectations model. However, PMD is not limited by linearity. PMD can be made more flexible in two different ways that do not significantly complicate the method. First, the first-step local projections can be made nonlinear in a variety of ways as described in Jordà (2005). Secondly, the method is not limited to linear rational expectations models. However, because the paper is already dense with results, we defer a more detailed discussion of these two issues to another paper. Finally, in addition to showing how PMD relates to GMM we also show how it relates to the Hannan-Rissanen (1982) estimator and how PMD can be used to estimate rather general time series models, including traditional ARMA(p,q) models.

2 Projection Minimum Distance: The Method

In this section we propose an estimator for rational expectations models (and more broadly time series models), where it is important to be flexible about the underlying dynamics of the data generating process (DGP) to ensure consistent estimates of the structural parameters of interest. The estimator we propose consists of two steps. In the first step we estimate impulse responses from the sample with a flexible, semi-parametric estimator: Jordà's (2005) local projections. In the second step we minimize the distance between the impulse responses estimated from the sample and the responses generated from the macroeconomic model being considered. This second step is based on Ferguson's (1958) minimum chi-square estimator. We call this two-step estimator "projection minimum distance" (PMD) and show that it is consistent and asymptotically normal.

We find it useful to use a generic rational expectations formulation (e.g., see Farmer, 1993; and Evans and Honkapohja, 2001) to illustrate the particulars of our technique. Hence, let $\mathbf{y}_t$ be an $r \times 1$ vector of endogenous variables whose behavior is described by the backward- and forward-looking model

$$\Phi_0' \mathbf{y}_t = \Phi_1' \mathbf{y}_{t-1} + \Phi_2' E_t \mathbf{y}_{t+1} + \mathbf{u}_t, \quad E(\mathbf{u}_t \mathbf{u}_t') = I \tag{1}$$

where $\mathbf{u}_t$ is the $r \times 1$ vector of expectational errors. The $r \times r$ coefficient matrix Φ_0 makes explicit the nature of the contemporaneous relations between elements of $\mathbf{y}_t$. Expression (1) is rather general. For instance, $\mathbf{y}_t$ can include exogenous variables whose economic processes are not directly modeled but whose reduced-form laws-of-motion can be expressed in finite Markov form. Similarly (1) does not restrict the model to have first order dynamics since the vector $\mathbf{y}_t$ can always be appropriately redefined to include lags of $\mathbf{y}_t$ and (1) can be thought of as a state-space representation. Expression (1) can therefore be appropriate for both dynamic stochastic general equilibrium or dynamic partial equilibrium models of the economy.

A stable solution for (1) is a dynamic stochastic difference equation in $\mathbf{y}_t$. In most practical situations, the stability of the solution implies that it is covariance-stationary, and hence, by the Wold decomposition theorem (see Anderson, 1994), can be represented as,

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j' \boldsymbol{\varepsilon}_{t-j}. \tag{2}$$

with $B_0 = I$, $E(\varepsilon_t \varepsilon_t') = \Omega$ and with the relation between the reduced form residuals of the Wold decomposition, ε_t, and the structural residuals of the economic model, $\mathbf{u}_t$, given by $\varepsilon_t = \Phi_0^{-1}\mathbf{u}_t$ (we omit the constant and deterministic terms for simplicity but without loss of generality). Substituting (2) into (1) and equating terms in ε_{t-j} (in analogous manner to the method of undetermined coefficients, see Christiano, 2002), we obtain the following set of conditions:

$$\begin{aligned} \Phi_0' &= \Phi_2' B_1' + \Phi_0' \quad \text{for } j = 0 \\ B_j' &= \Phi_1' B_{j-1}' + \Phi_2' B_{j+1} \quad \text{for } j \geq 1 \end{aligned} \tag{3}$$

where we remark that an estimate of Φ_0 can be obtained by imposing the structural identification conditions of the model and noticing that $\Omega = \Phi_0^{-1}\left(\Phi_0^{-1}\right)'$.

Momentarily ignoring the conditions associated with $j = 0$ in (3), it is clear that for $j \geq 1$, structural identification plays no role in estimating Φ_1 and Φ_2, which can be done on the basis of the reduced-form impulse responses, B_j, alone. Define

$$\underset{r(h+1)\times r}{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_h \end{bmatrix},$$

with $B_0 = I$; and the selector matrices

$$\begin{aligned} S_0 &= \begin{bmatrix} \underset{r(h-1)\times r}{\mathbf{0}} & (I_{h-1} \otimes I_r) & \underset{r(h-1)\times r}{\mathbf{0}} \end{bmatrix}, \\ S_1 &= \begin{bmatrix} (I_{h-1} \otimes I_r) & \underset{r(h-1)\times 2r}{\mathbf{0}} \end{bmatrix}, \\ S_2 &= \begin{bmatrix} \underset{r(h-1)\times 2r}{\mathbf{0}} & (I_{h-1} \otimes I_r) \end{bmatrix}. \end{aligned}$$

Then we can stack the conditions in (3) for $j = 1, ..., h$ more compactly as

$$S_0 B = S_1 B \Phi_1 + S_2 B \Phi_2 \tag{4}$$

The vec operator can be applied to both sides of this expression to vectorize the coefficient vectors. Let $b = vec(B)$; $\phi = \{vec(\Phi_1) \quad vec(\Phi_2)\}'$ then (4) can be written as

$$(I_r \otimes S_0)\, b = \{(I_r \otimes S_1 B) \quad (I_r \otimes S_2 B)\}\, \phi\,. \tag{5}$$

This last expression makes clear that if the impulse response coefficients in B were known rather than estimated, then ϕ could be simply estimated, say, by least squares. Instead, we show below that one can easily obtain semi-parametric estimates of b (or B interchangeably) with local projections in a first stage regression. In the second stage, we minimize the distance between the two sides of expression (5) by minimum chi-square methods.

In particular, let

$$g\left(\widehat{b}_T; \phi\right) = (I_r \otimes S_0)\, \widehat{b}_T - \left\{\left(I_r \otimes S_1 \widehat{B}_T\right) \quad \left(I_r \otimes S_2 \widehat{B}_T\right)\right\} \phi \tag{6}$$

then, a minimum chi-square estimator of ϕ can be found by minimizing

$$\min_{\phi} \widehat{Q}_T(\phi) = g\left(\widehat{b}_T;\phi\right)' \widehat{W} g\left(\widehat{b}_T;\phi\right) \tag{7}$$

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

Several remarks deserve comment. First, this estimation method is not limited by linear expectational mechanisms since (6) can be easily generalized appropriately. However, we do not explore the details of this feature explicitly in this paper. Second, the matrices Φ_1 and Φ_2 contain $2r^2$ parameters that we want to estimate but we have hr^2 conditions available for estimation, where h is determined by the sample size and the researcher in the same way that instruments are selected in GMM estimation. Thus, the availability of these overidentifying restrictions will provide a natural test of model misspecification. We show that this test is distributed chi-squared under standard regularity conditions that we discuss below. This test effectively measures the distance between the impulse responses from the data and those from the theoretical model. Macroeconomic models, whether calibrated, estimated by maximum likelihood, or estimated by some other method, are routinely evaluated by how well their impulse responses match those responses generated by the data. Hence, the overidentifying restrictions test provides a formal statistical metric for this common ocular diagnostic. Third, the relation between the coefficients in ϕ and $\widehat{b}_T$ is not a known function, as is commonly assumed in classical minimum distance estimation, but rather depends on the first-stage estimates $\widehat{b}_T$. Below we show how we incorporate this feature into the proofs of consistency and asymptotic normality.

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

3 First-Step: Local Projections

3.1 Consistency

In the previous section we argued that a stable solution of the rational expectations model (1) is therefore covariance-stationary and has a Wold decomposition,

$$\mathbf{y}_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j} \tag{8}$$

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

(i) $E(\varepsilon_t) = 0$ and ε_t are i.i.d.

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

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

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

Anderson (1994) shows that the process in (8) can also be written as:

$$\mathbf{y}_t = \sum_{j=1}^{\infty} A_j \mathbf{y}_{t-j} + \varepsilon_t \tag{9}$$

such that,

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

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

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

Results (iii) and (vi) imply that

$$\left[I_r - A_1 L - A_2 L^2 - ...\right] \left[I_r + B_1 L + B_2 L^2 + ...\right] = I_r$$

where L is the usual lag operator. Then, equating terms in the powers of L, we have the following correspondences:

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_1 B_1 + A_2 \\ &\vdots \\ B_h &= A_1 B_{h-1} + A_2 B_{h-2} + ... + A_{h-1} B_1 + A_h \end{aligned} \tag{10}$$

for any $h \geq 1$ with $B_0 = I_r$. Hence, expression (10) delivers a set of conditions that would allow us to determine the impulse response function for $\mathbf{y}_t$ if we knew what the A_j were. A truncated version of these conditions for a finite lag VAR is available in Hamilton (1994). In addition and by simple recursive substitution in the $VAR(\infty)$ representation, we have that

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + \varepsilon_{t+h} + B_1 \varepsilon_{t+h-1} + ... + B_{h-1} \varepsilon_{t+1} \tag{11}$$

where:

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

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

Now consider truncating the infinite lagged expression (11) at lag k

$$\mathbf{y}_{t+h} = A_1^h \mathbf{y}_t + A_2^h \mathbf{y}_{t-1} + ... + A_k^h \mathbf{y}_{t-k+1} + \mathbf{v}_{k,t+h} \tag{12}$$

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

In what follows, we show that least squares estimates of (12) produce consistent estimates for A_j^h for $j = 1, ..., k$. Notice that this is more than we need since we are only interested in the consistency of the A_1^h.

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

(i) $X_{t,k} = \left(\mathbf{y}'_t, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-k+1}\right)'$ that is, the regressors in (12).

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

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

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

$$\underset{r \times kr}{\widehat{A}}(k,h) = (\widehat{A}_1^h, ..., \widehat{A}_k^h) = \widehat{\Gamma}'_{1,k,h} \widehat{\Gamma}_k^{-1} \tag{13}$$

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

Theorem 1 ***Consistency.*** *Let $\{y_t\}$ satisfy (8) and assume that:*

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

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

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

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

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

Then:

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

The proof of this theorem is in the appendix and parallels the proof in Lewis and Reinsel (1985) on the consistency properties of estimates from a truncated VAR when the underlying model is a VAR(∞). A natural consequence of the theorem provides the essential result that we need, that local projection estimates are consistent for the impulse response coefficients, that is, $\widehat{A}_1^h \xrightarrow{p} B_h$.

3.2 Asymptotic Normality

We now show that least-squares estimates from the truncated projections in (12) are asymptotically normal, although for the purposes of the PMD estimator, proving that $\widehat{A}_1^h$ is asymptotically normally distributed would suffice. Notice that we can write

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

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

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

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

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

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

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

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

Define,

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

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

then

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

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

hence

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

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

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

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

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

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

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

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

Theorem 2 *Let $\{\mathbf{y}_t\}$ satisfy (8) and assume that*

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

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

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

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

Then

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

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

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

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

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

Theorem 3 *Let $\{\mathbf{y}_t\}$ satisfy (8) and assume*

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

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

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

Then

$$S_T \xrightarrow{d} N(0,\Omega_h)$$

The proof is provided in the appendix. Several results deserve comment. Observe that for each horizon h considered, the asymptotic variance-covariance matrix Ω_h is determined by the moving average structure of the residuals, $\Omega_h = \left(\Gamma_k^{-1}\otimes\Sigma_h\right)$; $\Sigma_h = \left(\Sigma_\varepsilon + \sum_{j=1}^{h-1}B_j\Sigma_\varepsilon B_j'\right)$. Jordà (2005) suggests that one way of estimating this variance-covariance matrix semiparametrically is with the Newey-West (1987) heteroskedasticity and autocorrelation consistent estimator. However, if the projections are estimated sequentially, notice that each of the projections up to horizon $h-1$ provide consistent estimates of B_j $j = 1,2,...,h-1$, which could then be used to estimate the asymptotic variance-covariance matrix parametrically. We postpone discussion on the differences between these two methods for a later paper and proceed with the alternative based on the Newey-West estimator.

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

(i) $\underset{r(k-2)\times 1}{X_{t-1,k}} \equiv \left(\mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-k+1}\right)'$

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

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

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

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

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

$$\widehat{\Gamma}'_{1,h|k}\widehat{\Gamma}^{-1}_{1|k} = \begin{bmatrix} \widehat{A}_1^1 \\ \widehat{A}_1^2 \\ \\ \widehat{A}_1^h \end{bmatrix} = \begin{bmatrix} \widehat{B}_1 \\ \widehat{B}_2 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = \widehat{B}(1,h) \tag{14}$$

Using the usual least-squares formulas, notice that

$$\widehat{B}(1,h) = B(1,h) + \left\{(T-k-h)^{-1}\sum_{t=k}^{T-h} \mathbf{y}_t M_{t-1,k}V'_{t,h}\right\}'\widehat{\Gamma}^{-1}_{1|k} + o_p(1) \tag{15}$$

where $V_{t,h} \equiv \left(\mathbf{v}'_{t+1}, ..., \mathbf{v}'_{t+h}\right)'$; $\mathbf{v}_{t+j} = \varepsilon_{t+j} + B_1\varepsilon_{t+j-1} + ... + B_{j-1}\varepsilon_{t+1}$ for $j = 1, ..., h$ and the terms vanishing in probability in (15) involve the terms U_{1T}, U_{2T}, and U_{3T} defined in the proof of theorem one, which makes use of the condition $k^{1/2}\sum_{j=k+1}^{\infty} ||A_j|| \rightarrow 0$ as $T, k \rightarrow \infty$. Under the conditions of theorem 2, wew can write

$$(T-k-h)^{1/2} vec\left(\widehat{B}(1,h) - B(1,h)\right) \xrightarrow{p} \tag{16}$$
$$(T-k-h)^{1/2}\, vec\left[\left\{(T-k-h)^{-1}\sum_{t=k}^{T-h} V_{t,h}M_{t-1,k}\mathbf{y}'_t\right\}\widehat{\Gamma}^{-1}_{1|k}\right]$$

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

Next notice that

$$(T-k-h)^{-1}\sum_{t=k}^{T-h} V_{t,h}V'_{t,h} \xrightarrow{p} \underset{rh\times rh}{\Sigma_v} \tag{17}$$

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

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

where $\varepsilon_{t,h} \equiv \left(\varepsilon'_{t+1}, ..., \varepsilon'_{t+h}\right)'$ and $\Psi_B = [\Psi_{B,1} \ldots \Psi_{B,h}]$ with

$$\Psi_{B,1} = \begin{bmatrix} I_r \\ B_1 \\ \vdots \\ B_{h-1} \end{bmatrix} \quad \Psi_{B,h} = \begin{bmatrix} 0_{r(h-1),r} \\ I_r \end{bmatrix} \text{ and for } 1 < \text{j} < \text{h}, \ \Psi_{B,j} = \begin{bmatrix} 0_{r(j-1),r} \\ I_r \\ \vdots \\ B_{j-1} \end{bmatrix}.$$

Then, assuming $\underset{r\times r}{\Sigma_\varepsilon} = E(\varepsilon_{t+1}\varepsilon'_{t+1})$, it is easy to show that

$$\Sigma_v = \Psi_B\{I_r \otimes \Sigma_\varepsilon\}\Psi'_B \tag{18}$$

and hence that

$$(T-k-h)^{1/2}\, vec\left(\widehat{B}(1,h) - B(1,h)\right) \xrightarrow{d} N\left(\mathbf{0}, \Omega_B\right)$$
$$\underset{r^2h\times r^2h}{\Omega_B} = \left(\underset{r\times r}{\Gamma_{1|k}^{-1}} \otimes \underset{rh\times rh}{\Sigma_v}\right)$$

In practice, one requires sample estimates $\widehat{\Gamma}_{1|k}^{-1}$ and $\widehat{\Sigma}_v$.We have already defined a sample estimate for the first term, however, the latter estimate deserves particular consideration. As we mentioned in our derivations for theorem 3, one avenue would be to estimate expression (18) by a Newey-West type estimator of the variance-covariance matrix Σ_v. However, given the parametric form of expression (18), it would be easy to construct a sample estimate of Ω_B by plugging-in the estimates $\widehat{B}(1,h)$ and $\widehat{\Sigma}_\varepsilon$ into the expression (18). Furthermore, feasible GLS estimates of $B(1,h)$ could be easily computed in two steps: (1) compute $\widehat{B}(1,h)$, $\widehat{\Sigma}_\varepsilon$, and hence $\widehat{\Sigma}_v$ (2) do GLS by scaling $Y_{t,h}$ with $\widehat{\Sigma}_v^{-1/2}$. We defer further investigation on the properties of this estimator to a different paper to avoid burdening the reader with too many results.

In practice, we find it convenient to estimate responses for horizons 1, ..., h jointly as follows. define $\mathbf{y}_j$ for $j = h$, ..., 1, 0, −1, ..., $-k$ as the $(T-k-h)\times r$ matrix of observations of the vector y_{t+j}. Additionally, define $Y \equiv (\mathbf{y}_1, ..., \mathbf{y}_h)'$; $X \equiv (I_h \otimes \mathbf{y}_0)$; $Z \equiv \left(\mathbf{1}_{(T-k-h)\times r}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k+1}\right)$ and hence, $M_z = I_{(T-k-h)h} - (I_h \otimes Z)\left[(I_h \otimes Z)'(I_h \otimes Z)\right]^{-1}(I_h \otimes Z)$. Standard properties of least-squares allow us to jointly estimate

$$\widehat{B}(1,h) = \begin{bmatrix} \widehat{B}_1 \\ \vdots \\ \widehat{B}_h \end{bmatrix} = \left[X'M_zX\right]^{-1}\left[X'M_zY\right] \tag{19}$$

and with asymptotic variance-covariance matrix for $vec(\widehat{B}(1,h))$, $\widehat{V}_b^2 = \left\{\left[X'M_zX\right]^{-1} \otimes \widehat{\Sigma}_h\right\}$; $\widehat{\Sigma}_h = \left(\widehat{\Sigma} + \sum_{j=1}^{h-1} \widehat{B}_j \widehat{\Sigma} \widehat{B}_j'\right)$, and $\widehat{\Sigma} = \frac{\widehat{\mathbf{v}}'\widehat{\mathbf{v}}}{T-k-h}$; $\widehat{\mathbf{v}} = M_zY - M_zX\widehat{B}(1,h)$. Hence, recall that what is needed for the two-step estimator is an estimate of $b = vec(B) = vec(B(0,h))$ where $B_0 = I_r$. Hence, from (19), $\widehat{b}_T = (vec(I_r)\ vec(\widehat{B}(1,h))'$ where $\widehat{V}_b^2$ can easily be redefined as

$$\widehat{V}_b^2 = \begin{pmatrix} \mathbf{0}_r & \mathbf{0}_{r\times(hr^2)} \\ \mathbf{0}_{(hr^2)\times r} & \widehat{V}_b^2 \end{pmatrix}$$

We conclude this section by highlighting its main results and drawing a brief comparison to estimates of the B_j based on a $VAR(k)$. First, our local projection estimator ensures that we have consistent estimates of B_j for $j = 1, ..., h$. We have shown that this result holds even when the underlying process is an infinite order VAR. Secondly, we have derived the variance-covariance matrix of $B(1,h)$. Finally, even if the underlying data generating process is more complex than an infinite order VAR, the nature of the local approximation of the projections is likely to provide more robust estimates of the B_j. In contrast, estimates of B_j based on a $VAR(k)$ have serious disadvantages. First, consistency of the B_j is only available for $j = 1, ..., k$. For any $j > k$ the B_j will be inconsistent for processes whose VAR representation has more than k lags (as would be expected with the $VARMA$ representations of the newest generation of dynamic, stochastic, general equilibrium model). Seconly, a closed form expression for the variance-covariance matrix of the B_j is unavailable. This prevents us from forming the optimal weighting matrix in the minimum distance step and complicates the computation of valid standard errors for the structural parameter estimates.

4 The Second Step: Minimum Chi-Square

4.1 Consistency

Given $\widehat{b}_T$ from the first-stage described in the previous sections, our objective here is to estimate $\phi = \{vec(\Phi_1)\ vec(\Phi_2)\}'$ from the conditions in (5) and minimization of expression (7), where expression (6) is rewritten here for convenience

$$g\left(\widehat{b}_T;\phi\right) = (I_r \otimes S_0)\,\widehat{b}_T - \left\{\left(I_r \otimes S_1 \widehat{B}_T\right) \quad \left(I_r \otimes S_2 \widehat{B}_T\right)\right\}\phi$$

and hence the quadratic function we need to minimize is

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

where $\widehat{\phi}_T$ denotes the sample estimate of ϕ from this two-step estimator and $Q_0(\phi)$ denotes the objective function at b_0. The following theorem establishes the conditions under which this estimator is consistent for ϕ_0.

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

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

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

(iii) The parameter space $\phi_0 \in \blacksquare$ *is compact*

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

(v) $\widehat{Q}_T(\phi)$ *converges uniformly in probability to* $Q_0(\phi)$

(vi) $hr^2 \geq \dim(\phi)$

Then

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

Proof is provided in the appendix. Next, we show that the minimum chi-square estimator is asymptotically normal.

4.2 Asymptotic Normality

The results in this section are based on Theorem 3.2 and pages 2149, 2175 in Newey and McFadden (1994). The following theorem derives the asymptotic distribution of $\widehat{\phi}_T$.

Theorem 5 *Assume:*

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

(ii) $\widehat{b}_T \xrightarrow{p} b_0$ *and* $\widehat{\phi}_T \xrightarrow{p} \phi_0$

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

(iv) $g(\widehat{b}_T; \phi)$ *is continuously differentiable in a neighborhood* $\mathfrak{N}$ *of* ϕ_0

(v) $\sqrt{T} g(b_0; \phi_0) = \sqrt{T} \left(I_r \otimes S_0\right) \left(\widehat{b}_T - b_0\right) \xrightarrow{d} N(0, V_B)$

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

$$\sup_{\phi \in \mathfrak{N}} \left\| \nabla_b g(\widehat{b}_T; \phi) - G_b \right\| \xrightarrow{p} 0$$
$$\sup_{\phi \in \mathfrak{N}} \left\| \nabla_\phi g(\widehat{b}_T; \phi) - G_\phi \right\| \xrightarrow{p} 0$$

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

(viii) $hr^2 \geq \dim(\phi)$

Then:

$$\sqrt{T}\left(\widehat{\phi}_T - \phi_0\right) \xrightarrow{d} N\left(0, V_\phi\right) \tag{20}$$
$$\begin{aligned} V_\phi = & \left(G'_\phi W G_\phi\right)^{-1} G'_\phi W \left(I \otimes S_0\right)' V_b \left(I \otimes S_0\right) W G_\phi \left(G'_\phi W G_\phi\right)^{-1} + \\ & \left(G'_\phi W G_\phi\right)^{-1} G'_b W \Sigma_h W G_b \left(G'_\phi W G_\phi\right)^{-1} + \\ & \left(G'_\phi W G_\phi\right)^{-1} \{G'_\phi W G_b - G'_b M_z X\}' V_b \{G'_\phi W G_b - G'_b M_z X\} \left(G'_\phi W G_\phi\right)^{-1} \end{aligned}$$

The optimal weighting matrix W is

$$W_o = \left\{\left(I \otimes S_0\right)' V_b \left(I \otimes S_0\right)\right\}^{-1}$$

which would simplify the expression of the variance further into

$$V_\phi = \left(G'_\phi W_o G_\phi\right)^{-1} + V^o_{\phi(\widehat{b}_T)}$$

where $V^o_{\phi(\widehat{b}_T)}$ denotes that portion of the variance of $\widehat{\phi}_T$ due to the fact that $\widehat{b}_T$ is estimated and corresponds to the last two lines in expression (20) evaluated at W_o. Absent this uncertainty, then the expression for V_ϕ would simplify to $\left(G'_\phi W_o G_\phi\right)^{-1}$, the usual minimum distance result with optimal weighting matrix.

5 PMD: A Summary for Practitioners

The PMD estimator of the model in expression (1) and repeated here for convenience

$$\Phi_0' \mathbf{y}_t = \Phi_1' \mathbf{y}_{t-1} + \Phi_2' E_t \mathbf{y}_{t+1} + \mathbf{u}_t, \quad E\left(\mathbf{u}_t' \mathbf{u}_t\right) = I$$

consists of the following steps:

1. Construct $Y = (\mathbf{y}_1, ..., \mathbf{y}_h)'$; $X = (I_h \otimes \mathbf{y}_0)$; $Z = \left(\mathbf{1}_{(T-k-h)\times r}, \mathbf{y}_{-1}, ..., \mathbf{y}_{-k+1}\right)$; $M_z = I_{(T-k-h)h} - (I_h \otimes Z)\left[(I_h \otimes Z)'(I_h \otimes Z)\right]^{-1}(I_h \otimes Z)$, where $\mathbf{y}_j$ is the $(T-k-h)\times r$ matrix of observations for the vector $\mathbf{y}_{t+j}$.

2. Compute by least squares $\widehat{b}_T = vec(I_r \quad \widehat{B}(1,h))$, where

$$\widehat{B}(1,h) = \left[X' M_z X\right]^{-1}\left[X' M_z Y\right]$$

3. Compute $\widehat{V}_b$ as $\widehat{V}_b = (T-k-h)^{-1}\left\{\left[X' M_z X\right]^{-1} \otimes \widehat{\Sigma}_h\right\}$; $\widehat{\Sigma}_h$ estimated by Newey-West on the residuals of the regression in bullet point 2. Remember to redefine $\widehat{V}_b$ by adding the appropriate zeros corresponding to the variance and covariances of the term I in $\widehat{b}_T$

4. Define $\widehat{Y}_b = (I_r \otimes S_0)\,\widehat{b}_T$; $\widehat{X}_b = \left\{\left(I_r \otimes S_1 \widehat{B}_T\right) \quad \left(I_r \otimes S_2 \widehat{B}_T\right)\right\}$ where S_0, S_1, S_2 are the matrices that select the appropriate elements of $\widehat{B}_T = \widehat{B}(1,h)$. Further define $\phi = \{vec(\Phi_1) \quad vec(\Phi_2)\}$.

5. The minimum chi-square estimate of ϕ is easily computed by weighted least-squares

$$\widehat{\phi}_T = \left(\widehat{X}_b \widehat{V}_b^{-1} \widehat{X}_b\right)^{-1}\left(\widehat{X}_b \widehat{V}_b^{-1} \widehat{Y}_b\right)$$

6. The variance-covariance matrix for $\widehat{\phi}_T$ can be calculated as:

$$\widehat{V}_\phi = \frac{1}{T-k-h}\left(\widehat{X}_b' \widehat{W}_o \widehat{X}_b\right)^{-1} + \frac{1}{T-k-h}\widehat{V}^o_{\phi\left(\widehat{b}_T\right)}$$

where $\widehat{W}_o = \left\{(I \otimes S_0)'\,\widehat{V}_b\,(I \otimes S_0)\right\}^{-1}$ and the expression for $\widehat{V}^o_{\phi\left(\widehat{b}_T\right)}$ can be computed directly from expression (20).

7. A test of misspecification can be constructed by noticing that

$$\left(\widehat{Y}_b - \widehat{X}_b \widehat{\phi}_T\right)' \widehat{W}_0 \left(\widehat{Y}_b - \widehat{X}_b \widehat{\phi}_T\right) \xrightarrow{d} \chi^2_{(h-2)r^2}$$

6 Invalid Instruments: PMD versus GMM

We begin this section with a simple motivating example. Suppose the DGP is characterized by the backward/forward model:

$$y_t = \phi_1 y_{t-1} + \phi_2 E_t y_{t+1} + \varepsilon_t. \tag{21}$$

Instead, Euler conditions of conventional rational expectations models can be expressed as

$$y_t = \rho E_t y_{t+1} + u_t, \tag{22}$$

which in this example, are misspecified with respect to the DGP. Based on the economic model in (22), y_{t-j}; $j > 1$ would be considered valid instruments for GMM estimation and hence, an estimate of ρ would be found with the set of conditions

$$\widehat{\rho}_{GMM} = \left(\frac{1}{T} \sum^{T} y_{t+1} y_{t-j} \right)^{-1} \left(\frac{1}{T} \sum^{T} y_{t-j} y_t \right). \tag{23}$$

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

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

where $\gamma_j = COV(y_t y_{t-j})$. Notice that the bias, $\phi_1 \frac{\gamma_{j-1}}{\gamma_{j+1}}$, does not disappear by selecting longer lags of y_{t-j} as instruments, the usual GMM recommendation. In fact, any y_{t-j}; $j \geq 1$ is an invalid instrument and although $\gamma_j \to 0$ as $j \to \infty$, $\frac{\gamma_{j-1}}{\gamma_{j+1}}$ is indeterminate, since both covariances are simultaneously going to zero. Meanwhile, as $j \to \infty$ the correlation of the instrument with the regressor is exponentially decaying to zero.

Instead consider what happens in PMD. Notice that the MA(∞) representation of (21) is

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

and hence, under the proposed model in (22), we obtain the following set of conditions analogous to (3), namely

$$\begin{bmatrix} b_1 \\ \vdots \\ b_h \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ b_{h-1} \end{bmatrix} \rho \tag{24}$$

which we can use to set-up the PMD estimator. Local projections for the b_j are

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

where $M_t = 1 - z_t \left(z'_t z_t \right)^{-1} z_t$ and $z_t = (1 \; y_{t-2} \; ... y_{t-k+1})$. In order to compare PMD directly with the GMM expression (23) notice that

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

then, for $j \geq 1$

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

and hence

$$\widehat{\rho}_{PMD} \xrightarrow{p} \phi_1$$

since

$$\frac{1}{T} \sum^{T} y_{t-j-1} M_{t-j} y_t \xrightarrow{p} 0; \frac{1}{T} \sum^{T} \varepsilon_{t-j} M_{t-j} y_t \xrightarrow{p} 0.$$

Clearly, a main distinguishing characteristic between PMD and GMM is that PMD uses the semiparametric estimates of the impulse responses to generate conditional instruments that are valid even if the model is misspecified. In contrast, GMM uses endogenous instruments unconditionally and therefore their validity heavily depends on having specified the correct model. A second advantage of PMD is that the optimal weights assigned to each impulse responses condition are inversely proportional to the estimated variance of the impulse response coefficient. Hence, it is the marginal informational content of this conditional instrument that determines its contribution to the parameter estimates and its variance.

7 Monte Carlo Experiments (incomplete)

7.1 Estimating ARMA(p,q) models with PMD

This section investigates the small sample properties of PMD. We take this opportunity to further demonstrate the flexibility of our method by experimenting with univariate ARMA specifications, which would typically require numerical optimization routines. In particular, consider a covariance-stationary ARMA(p,q) model,

$$y_t = \rho_1 y_{t-1} + ... + \rho_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \tag{25}$$

with Wold decomposition,

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \tag{26}$$

with $b_0 = 1$. Matching the responses b_i in (26) into (25) delivers the set of conditions

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_h \end{bmatrix} = \begin{bmatrix} 1 \\ b_1 \\ b_2 \\ \vdots \\ b_{h-1} \end{bmatrix} \rho_1 + \begin{bmatrix} 0 \\ 1 \\ b_1 \\ \vdots \\ b_{h-2} \end{bmatrix} \rho_2 + ... + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ b_{h-p} \end{bmatrix} \rho_p + \tag{27}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \theta_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \theta_2 + ... + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \theta_q$$

Table 1 summarizes a battery of experiments based on several ARMA(1,1) specifications that include pure AR and pure MA examples as well. For each choice of ARMA parameters indicated in the table, residuals are drawn from a standard normal and used to generate samples of 100 and 300 observations, where the first 100 observations generated are discarded to avoid initialization problems. The horizon h is chosen when a joint F-test of the $h+1$ projection fails to reject the null that the projection coefficients are zero. Finally, we compare the properties of ERME against the Durbin-Hannan-Rissanen estimator for ARMA models (see Durbin, 1960 and Hannan-Rissanen, 1982). This is a two-step estimator based on fitting a long autoregression in the first step and then using the residuals to estimate the ARMA model in the second step. We choose this method for several reasons. Maximum likelihood estimation requires numerical routines which crashed for some draws of the residuals. The Durbin-Hannan-Rissanen estimator does not have these numerical problems and provides asymptotically consistent and efficient estimates[1] .

Overall, we found the results of our Monte-Carlo experiment very encouraging. ERME esti-

[1] The GAUSS routines used to estimate the models with the Durbin-Hannan-Rissanen approach were obtained from the collection of time series routines prepared by Rainer Schlittgen and Thomas Noack and available at the American University repository: http://www.american.edu/academic.depts/cas/econ/gaussres/timeseri/timeseri.htm

mates are numerically very similar (even slightly more accurate) to the Durbin-Hannan-Rissanen estimates and close to the true parameter values. Although not reported in the table, we found the numerical estimates and associated standard errors to be very stable to the choice of truncation horizon h. Standard errors from ERME tended to be slightly less efficient, the difference becoming smaller for larger samples or when the truncation horizon is larger.

7.2 Fuhrer and Olivei (2004)

7.3 Systems

8 Application: Fuhrer and Olivei (2004) revisited

The popular, elementary New-Keynesian framework for monetary policy analysis combines forward-looking, micro-founded output (IS curve) and inflation (AS curve) Euler equations with a policy reaction function. The specific expectations mechanism considered plays a prominent role: purely forward-looking expectations mechanisms seem to be at odds with the dynamic properties of the data. Fuhrer and Olivei (2004) investigate whether the nature of this dissonance can be explained by the poor properties of weak instruments GMM or whether "hybrid" specifications, that simultaneously incorporate forward and backward-looking behavior, are warranted. To this end, Fuhrer and Olivei (2004) propose a GMM procedure that imposes the dynamic constraints implied by the economic model on the instruments. They dub this procedure "optimal-instruments" GMM (OI-GMM) and explore its properties relative to conventional GMM and MLE.

This section uses ERME on the specifications in Fuhrer and Olivei (2004) to offer a benchmark comparison between GMM, MLE and OI-GMM, and ERME. The basic specification is:

$$z_t = (1-\mu)\, z_{t-1} + \mu E_t z_{t+1} + \gamma E_t x_t + \varepsilon_t \tag{28}$$

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 (28), z_t in a measure of inflation, x_t is a measure of the output gap, and $\gamma > 0$ signifying that a positive output gap exerts "demand pressure" on inflation.

Fuhrer and Olivei (2004) experiment with a quarterly sample from 1966:Q1 to 2001:Q4 and use the following measures for z_t and x_t. The output gap is measured, either by the log deviation of real GDP from its HP trend or, from a segmented time trend with breaks in 1974 and 1995. Real interest rates are measured by the difference of the federal funds rate and next period's inflation. Inflation is measured by the log change in the GDP, chain-weighted price index. In addition, Fuhrer and Olivei (2004) experiment with real unit labor costs instead of the output gap for the inflation Euler equation. Further details can be found in their paper. Accordingly, Tables 2 and 3 reproduce Tables 4 and 5 in Fuhrer and Olivei (2004) and incorporate the results we find from ERME. In addition, we report unconstrained estimates of expression (28) labeled with the superscript "u" for unconstrained ("c" for constrained). The projection lag length of ERME's first step is determined by Akaike's information criterion and the horizon truncation h is determined with a joint significance F-test of the projections at each horizon, as we did in the Monte Carlos.

Table 2 shows significant discrepancies between GMM, MLE and OI-GMM on one hand, and ERME on the other. Where as the former methods' estimates of μ range between 0.52 to 0.41, ERME's estimates are in the range -0.40 to -0.17. Interestingly, ERME finds a much stronger link and of the correct sign between current and future interest rates and the output gap - a coefficient that is routinely estimated to be insignificant in the literature. Our estimates range from -0.13 to -0.10, whereas maximum likelihood suggests this coefficient to be -0.0084 at best. At worst the coefficient is of the wrong sign and insignificant in all other cases.

We investigated two possible explanations of the discrepancies between methods. First, we considered whether the constraint that the coefficient on z_{t-1} and the coefficient on $E_t z_{t+1}$ add up to one and imposed by Fuhrer and Olivei (2004), was being violated by the data. This turned out not to be the case as can be observed from Table 2. Secondly, Cameron and Trivedi (2005) suggest that in small samples, optimally weighted minimum distance estimators can be biased and that it may be preferable to use a weighting matrix with equal weights. The results with equal weights ERME are reported in the bottom third of table 2. Estimates of μ and γ under both the constrained and unconstrained versions changed very little indeed, suggesting that this was not the source of discrepancy.

What could explain the differences then? The overall specification J-test from ERME suggests that, whether unconstrained or not, estimated with equal weights or not, the theoretical model is misspecified. Because MLE and OI-GMM impose this theoretical and misspecified model on the data, we attribute differences across methods to this constraint. This impression was further reinforced by the results reported in Table 3. As we shall see momentarily, when the J-test indicates that the theoretical model is not misspecified, estimates from ERME are almost identical to MLE and OI-GMM. Because ERME is robust to misspecification of the underlying data generating process, we feel strongly about our results if only the controlled environment of a Monte-Carlo exercise could arbitrate among the merit of each method.

Table 3 reports estimates of the inflation Euler equation. As we remarked above, when ERME's J-test does not find evidence of misspecification, ERME's estimates are very close to the MLE estimates (in fact, somewhat closer than OI-GMM). μ is estimated to be approximately 0.20 and the output gap enters with the correct sign and with a coefficient of about 0.10. However, discrepancies among methods arise when we use real unit labor costs instead of the output gap. In this case, the J-test suggests the model is misspecified and, while ERME finds a value of μ which is still in the neighborhood of 0.20, it finds that the effect of real unit labor costs on inflation is much larger: 0.35 and highly significant. In contrast, both ML and OI-GMM have μ growing from 0.20 to 0.45 while the effect of real unit labor costs becomes insignificant.

We conclude this section by observing that ERME estimates were insensitive to the choice of projection lag length (we obtained similar results with varying lag length specifications) and changed very little when we vary the number of impulse response horizons used. Obviously, while our method produces results that are preferable from the economic point of view (in that the coefficient estimates of γ are more closely in line with a priori economic intuition), the more sensible argument comes from observing that, while MLE and OI-GMM impose the economic model on the data (and hence their properties cannot be assessed when this assumption is violated), ERME does no such thing – it uses local projections to obtain the most accurate description of the properties of the data generating process and uses these to obtain estimates of the parameters of interest.

9 Conclusions

This paper introduces a disarmingly simple method to estimate rational expectations models based on mapping the structural parameters of the economic model into the impulse responses from the data. The method avoids many of the computational complexities associated with full-information maximum likelihood or the optimal instruments GMM approach proposed by Fuhrer and Olivei (2004), yet possesses several advantages over these alternative estimation procedures. By assuming that the solution paths and exogenous stochastic processes in the model are characterized by a Wold representation rather than by a specific Markov process, we require less knowledge about the data generating process and can accommodate more general dynamic specifications. Perhaps most importantly, our efficient response matching estimator is designed to estimate parameters so that the model fits the data precisely along the dimensions against which it will be evaluated. The success or failure of dynamic stochastic general equilibrium models is ultimately based on their ability to explain the dynamic interactions among the economic variables involved, which interactions are typically summarized by impulse response functions. It is in this sense that our method conforms to the introductory quote by Sims.

An important feature of our method originates from how we adapt the local projection response estimator proposed by Jordà (2005). This impulse response estimator is more robust to general misspecification of the relevant underlying stochastic processes. This feature is particularly advantageous as it allows the ERME method to extract additional information from longer horizon impulse responses in cases where the Wold decomposition differs from the typically assumed Markov format. In the empirical example, this additional information tightened estimates of standard errors considerably.

New second and higher order accurate solution techniques for nonlinear dynamic stochastic general equilibrium models (see, e.g. Kim et al., 2003) deliver equilibrium conditions that are polynomial (rather than linear) difference equations. Future research will investigate the ERME method in these environments. We expect that, while the relationship between impulse response coefficients and the model's parameters will no longer be linear, the only real variation in our method will be a nonlinear-GLS type step, which is still relatively straight-forward to apply and may open the door to formal evaluation of this next generation of models.

10 Appendix

Proof. Theorem 1

Notice that

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

where

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

Hence,

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

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

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

Hence

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

where

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

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

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

by independence. Hence

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

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

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

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

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

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

Hence

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

by recursive substitution. Thus

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

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

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

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

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

instead. ■

Proof. Theorem 2

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

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

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

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

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

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

and substituting this result into their proof. ■

Proof. Theorem 3

Follows directly from Lewis and Reinsel (1985) by redefining

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

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

Proof. Theorem 4

Notice that when $\widehat{b}_T \xrightarrow{p} b_0$, then

$$\begin{aligned} g(\widehat{b}_T; \phi) &= (I_r \otimes S_0)\, \widehat{b}_T - \left\{ \left(I_r \otimes S_1 \widehat{B}_T \right) \quad \left(I_r \otimes S_2 \widehat{B}_T \right) \right\} \phi \xrightarrow{p} \\ &= (I_r \otimes S_0)\, b_0 - \{ (I_r \otimes S_1 B_0) \quad (I_r \otimes S_2 B_0) \} \phi \\ &= g\,(b_0; \phi) \end{aligned}$$

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

$$Q_0(\phi) = g(b_0;\phi)' W g(b_0;\phi)$$

which is a quadratic expression that is uniquely maximized at ϕ_0. $Q_0(\phi)$ is obviously a continuous function. Hence, the last thing we need to show is that

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

uniformly. [I have not done this formally yet. Here is where the stochastic equicontinuity may come in] ■

Proof. Theorem 5

The strategy of the proof consists in stacking the first and second steps into a vector of conditions whose weighted quadratic distance we want to minimize. This set-up then allows us to apply the mean value theorem on the set of first order conditions and derive the desired result. Therefore, let

$$m_T(b) = M_z Y - M_z X b$$

and notice that

$$\min_b m_T(b)' m_T(b)$$

results in

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

the local projection estimator for the first-step, reduced-form, impulse responses we presented in section 3. Hence, stacking the first and second stage conditions into a vector, we have

$$R_T(b, \phi) = \begin{bmatrix} m_T(b)' & g\left(\widehat{b}_T; \phi\right)' \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & W \end{bmatrix} \begin{bmatrix} m_T(b) \\ g\left(\widehat{b}_T; \phi\right) \end{bmatrix}.$$

The first order conditions of the problem

$$\min_{\beta, \phi} R_T(b, \phi)$$

are:

$$\begin{bmatrix} \nabla_b m_T(b)' & 0 \\ \nabla_b g\left(\widehat{b}_T; \phi\right)' & \nabla_\phi g\left(\widehat{b}_T; \phi\right)' \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & W \end{bmatrix} \begin{bmatrix} m_T(b) \\ g\left(\widehat{b}_T; \phi\right) \end{bmatrix} = 0. \qquad (29)$$

Notice that

$$
\begin{aligned}
\nabla_b m_T(b) &= -M_z X \\
\nabla_b g(b;\phi) &= (I \otimes S_0) - \{(\Phi_1' \otimes I)(I \otimes S_{-1}) + (\Phi_2' \otimes I)(I \otimes S_1)\} \\
\nabla_\phi g(b;\phi) &= -\{(I \otimes S_{-1}) B \quad (I \otimes S_1) B\}
\end{aligned}
$$

We find it convenient to define $G_b = \nabla_b g(b;\phi)$ and $G_\phi = \nabla_\phi g(b;\phi)$ and we use $\overline{G}_b = \nabla_b g\left(\overline{b};\overline{\phi}\right)$ to evaluate the gradient at a mean value $\overline{b} \in \left[\widehat{b}_T, b_0\right]$ and $\widehat{G}_b = \nabla_b g\left(\widehat{b}_T;\widehat{\phi}_T\right)$ and similarly for $\overline{G}_\phi$ and $\widehat{G}_\phi$.

Next, apply the mean value theorem to the last term in the first order conditions (29):

$$
\begin{bmatrix} m_T(b) \\ g\left(\widehat{b}_T;\phi\right) \end{bmatrix} = \begin{bmatrix} m(b_0) \\ g(b_0;\phi_0) \end{bmatrix} + \begin{bmatrix} -M_z X & 0 \\ \overline{G}_b & \overline{G}_\phi \end{bmatrix} \begin{bmatrix} \widehat{b}_T - b_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \tag{30}
$$

and plug this expression back into the first order conditions (29)

$$
\begin{bmatrix} -X'M_z & 0 \\ \widehat{G}_b' & \widehat{G}_\phi \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & W \end{bmatrix} \times
$$

$$
\left\{ \begin{bmatrix} m(b_0) \\ g(b_0;\phi_0) \end{bmatrix} + \begin{bmatrix} -M_z X & 0 \\ \overline{G}_b & \overline{G}_\phi \end{bmatrix} \begin{bmatrix} \widehat{b}_T - b_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} \right\} = 0
$$

After tedious algebra, we obtain

$$
\begin{bmatrix} \widehat{b}_T - b_0 \\ \widehat{\phi}_T - \phi_0 \end{bmatrix} = \tag{31}
$$

$$
\begin{bmatrix} (X'M_z Z)^{-1} & 0 \\ -\left(\widehat{G}_\phi' W \overline{G}_\phi\right)^{-1} \left\{\widehat{G}_\phi' W \overline{G}_b - \widehat{G}_b' M_z X\right\} (X'M_z Z)^{-1} & \left(\widehat{G}_\phi' W \overline{G}_\phi\right)^{-1} \end{bmatrix} \times
$$

$$
\begin{bmatrix} -X'M_z & 0 \\ \widehat{G}_b' & \widehat{G}_\phi' W \end{bmatrix} \begin{bmatrix} m(b_0) \\ g(b_0;\phi_0) \end{bmatrix}
$$

Notice that from conditions and assumptions of the problem,

$$\sqrt{T}m\left(b_0\right) \xrightarrow{d} N\left(0,\Sigma_h\right)$$

$$\sqrt{T}\left(I\otimes S_0\right)g\left(b_0;\phi_0\right) \xrightarrow{d} N\left(0,V_b\right)$$

$$\widehat{G}_b,\overline{G}_b \xrightarrow{p} G_b$$

$$\widehat{G}_\phi,\overline{G}_\phi \xrightarrow{p} G_\phi$$

Therefore, the second row of expression (31) allows us to write

$$\begin{aligned}
&\sqrt{T}\left(\widehat{\phi}_T-\phi_0\right) \xrightarrow{d} N\left(0,V_\phi\right)\\
V_\phi \;=\;& \left(G'_\phi W G_\phi\right)^{-1} G'_\phi W\left(I\otimes S_0\right)' V_b\left(I\otimes S_0\right) W G_\phi\left(G'_\phi W G_\phi\right)^{-1}+\\
&\left(G'_\phi W G_\phi\right)^{-1} G'_b W \Sigma_h W G_b\left(G'_\phi W G_\phi\right)^{-1}+\\
&\left(G'_\phi W G_\phi\right)^{-1}\{G'_\phi W G_b-G'_b M_z X\}' V_b\{G'_\phi W G_b-G'_b M_z X\}\left(G'_\phi W G_\phi\right)^{-1}
\end{aligned}$$

where we make use of the fact that

$$\left(X'M_zX\right)^{-1}X'M_z\Sigma_hM_zX\left(X'M_zX\right)^{-1}=V_b$$

■

References

Blanchard, Olivier and Danny Quah (1989) "The Dynamic Effects of Aggregate Demand and Aggregate Supply Disturbances," *American Economic Review*, 79: 655-73

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

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

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*, v. 113, 1-45.

Cooley, Thomas F. (1997) "Calibrated Models," *Oxford Review of Economic Policy,* 13:55-69.

Durbin, James (1960) "The Fitting of Time-Series Models," *Revue de l'Institut International de Statistique*, 28(3), 233-244.

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

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

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

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

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

Kim, Jinill, Sunghyun Kim, Ernst Schaumburg, and Christopher A. Sims (2003) "Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models," Federal Reserve Board, Finance and Economics Discussion Series, 2003-61.

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

Harvey, James E., Kevin D. Hoover and Kevin D. Salyer, eds. (1998) **Real Business Cycles, A Reader**. London and New York: Routledge.

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

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

McCallum, Bennett T. (1983) "On Non-Uniqueness in Rational Expectations Models: An Attempt at Perspective," *Journal of Monetary Economics*, March, 139-168.

McCallum, Bennett T. (2003) "The Unique Minimum State Variable RE Solution is E-Stable in All Well Formulated Linear Models," National Bureau of Economic Research, working paper 9960.

Rotemberg, Julio J. and Michael Woodford (1997) "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in **NBER Macroeconomics Annual 1997**, Julio J. Rotemberg and Ben S. Bernanke, eds. Cambridge and London: MIT Press.

Rudebusch, Glenn D. and Jeffrey C. Fuhrer (2004) "Estimating the Euler Equation for Output," *Journal of Monetary Economics*, 51(6): 1133-53.

Sbordone, Argia M. (2004) "A Limited Information Approach to the Simultaneous Estimation of Wage and Price Dynamics," Rutgers University, *mimeo.*

Sbordone, Argia M. (2005) "Do Expected Future Marginal Costs Drive Inflation Dynamics?" *Federal Reserve Bank of New York*, Staff Report n0. 204.

Sims, Christopher A. (1996) "Macroeconomics and Methodology," *Journal of Economic Perspectives,* v. 10, n.1, 105-120.

Stock, James H., Jonathan 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, 518-529.

Watson, Mark (1993) "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101:1011-1041.

Table 1. Fitting ARMA(p,q) Models with Efficient Response Matching Estimation. A Monte-Carlo Comparison

Data simulated from the model:

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

	Model				
	$\rho = 0.25$ $\theta = 0.50$	$\rho = 0.50$ $\theta = 0.25$	$\rho = 0.50$ $\theta = 0.50$	$\rho = 0.75$ $\theta = 0$	$\rho = 0$ $\theta = -0.50$
T = 100					
DHR	0.264	0.479	0.483	0.702	-0.034
	(0.146)	(0.135)	(0.110)	(0.099)	(0.190)
	0.482	0.264	0.518	0.019	-0.482
	(0.130)	(0.147)	(0.107)	(0.138)	(0.167)
ERME*	0.268	0.461	0.479	0.716	-0.019
	(0.152)	(0.143)	(0.123)	(0.133)	(0.220)
	0.472	0.259	0.500	0.010	-0.507
	(0.177)	(0.166)	(0.152)	(0.165)	(0.240)
h	5	4	5	4	3
T = 300					
DHR	0.246	0.485	0.482	0.747	-0.006
	(0.084)	(0.076)	(0.064)	(0.051)	(0.112)
	0.495	0.251	0.507	-0.005	-0.501
	(0.075)	(0.084)	(0.062)	(0.078)	(0.097)
ERME*	0.246	0.482	0.483	0.748	-0.020
	(0.092)	(0.084)	(0.071)	(0.061)	(0.124)
	0.494	0.263	0.501	-0.005	-0.500
	(0.107)	(0.101)	(0.090)	(0.074)	(0.136)
h	7	8	9	12	5

Notes: DHR refers to the Durbin-Hannan-Rissanen estimator for ARMA models. ERME* refers to the ERME estimates obtained from choosing the maximum impulse response horizon *h* for which the joint null that all the coefficients are zero is rejected at a 95% confidence level. This choice of horizon is reported in the row labeled *h*. *T = 100, 300* refers to the sample size. The first 100 observations of each simulation are discarded to avoid initialization problems. The Monte-Carlo is replicated 200 times. The numbers reported are the Monte-Carlo averages of the estimates (to assess consistency) and the standard errors are the Monte-Carlo average over the standard errors obtained from each of the two estimation alternatives (to assess efficiency).

Table 2 – ERME, MLE, GMM and Optimal Instruments GMM: A Comparison

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

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

Estimation Method	Specification	μ SE(μ)	γ SE(γ)	p-value J-Statistic	1st Step Lag Length
GMM	HP	0.52 (0.053)	0.0024 (0.0094)		
GMM	ST	0.51 (0.049)	0.0029 (0.0093)		
ML	HP	0.47 (0.035)	-0.0056 (0.0037)		
ML	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)		
Optimally Weighted ERME (optimal horizon, h = 4)					
$ERME^c$	HP	-0.40 (0.067)	-0.10 (0.075)	0.015	4
$ERME^c$	ST	-0.32 (0.062)	-0.14 (0.075)	0.003	4
$ERME^u$	HP	1.43/-0.41 (0.12/0.07)	-0.10 (0.075)	0.016	4
$ERME^u$	ST	1.45/-0.32 (0.11/0.06)	-0.13 (0.075)	0.007	4
Equally Weighted ERME (optimal horizon, h = 4)					
$ERME^c$	HP	-0.23 (0.123)	-0.10 (0.082)	0.001	4
$ERME^c$	ST	-0.17 (0.119)	-0.10 (0.086)	0.000	4
$ERME^u$	HP	1.18/-0.25 (0.17/0.12)	-0.09 (0.101)	0.001	4
$ERME^u$	ST	1.21/-0.17 (0.16/0.12)	-0.11 (0.10)	0.000	4

Notes: Top third of the table is Table 4 in Fuhrer and Olivei (2004). Estimates based on 4 horizons of the impulse response (determined by a joint F-test on the projections) and projection lag-length is determined by AICc. HP is the Hodrick-Prescott filter of log real GDP, and ST is a segmented deterministic trend of log real GDP. $ERME^c$ constrains the coefficients on z_{t-1} and $E_t z_{t+1}$ to add up to one. $ERME^u$ keeps the coefficients unconstrained and reports the value of the coefficient on z_{t-1} and then $E_t z_{t+1}$. Optimal Weights ERME uses the variance-covariance matrix from the projection step as weights in the second step. Equally weighted ERME uses the identity matrix instead. P-value of the J-Statistic is an overall specification test. Hence rejection of the null (i.e., p-value < 0.05, say), indicates model misspecification.

Table 3 – ERME, MLE, GMM and Optimal Instruments GMM: A Comparison Estimates of Inflation Euler Equation: 1966:Q1 to 2001:Q4

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

Estimation Method	Specification	μ SE(μ)	γ SE(γ)	p-value J-Statistic	1st Step Lag Length
GMM	HP	0.66	-0.055		
		(0.13)	(0.072)		
GMM	ST	0.63	-0.030		
		(0.13)	(0.050)		
GMM	*rulc*	0.60	0.053		
		(0.086)	(0.038)		
ML	HP	0.17	0.10		
		(0.037)	(0.042)		
ML	ST	0.18	0.074		
		(0.036)	(0.034)		
ML	*rulc*	0.47	0.050		
		(0.024)	(0.0081)		
OI – GMM	HP	0.23	0.12		
		(0.093)	(0.042)		
OI – GMM	ST	0.21	0.097		
		(0.11)	(0.039)		
OI - GMM	*rulc*	0.45	0.054		
		(0.028)	(0.0081)		
Optimally Weighted ERME (optimal horizon, h = 4)					
$ERME^c$	HP	0.21	0.09	0.714	3
		(0.11)	(0.06)		
$ERME^c$	ST	0.21	0.07	0.829	3
		(0.11)	(0.06)		
$ERME^c$	*rulc*	0.22	0.32	0.000	2
		(0.10)	(0.07)		
$ERME^u$	HP	0.59/0.22	0.09	0.912	3
		(0.19/0.11)	(0.06)		
$ERME^u$	ST	0.61/0.22	0.07	0.963	3
		(0.19/0.11)	(0.06)		
$ERME^u$	*rulc*	0.56/0.18	0.35	0.003	2
		(0.14/0.10)	(0.08)		
Equally Weighted ERME (optimal horizon, h = 4)					
$ERME^c$	HP	0.21	0.09	0.432	3
		(0.11)	(0.06)		
$ERME^c$	ST	0.21	0.07	0.829	3
		(0.11)	(0.06)		
$ERME^c$	*rulc*	0.22	0.32	0.000	2
		(0.10)	(0.07)		
$ERME^u$	HP	0.59/0.22	0.09	0.913	3
		(0.19/0.11)	(0.06)		
$ERME^u$	ST	0.61/0.22	0.07	0.963	3
		(0.19/0.11)	(0.06)		
$ERME^u$	*rulc*	0.56/0.18	0.35	0.003	2
		(0.14/0.10)	(0.08)		

Notes: Top third of the table is Table 5 in Fuhrer and Olivei (2004). Estimates based on 4 horizons of the impulse response (determined by a joint F-test on the projections) and projection lag-length is determined by AICc. HP is the Hodrick-Prescott filter of log real GDP, and ST is a segmented deterministic trend of log real GDP. When the entry *rulc* appears, the specification replaces the output gap with real unit labor costs as the driving process. $ERME^c$ constrains the coefficients on z_{t-1} and $E_t z_{t+1}$ to add up to one. $ERME^u$ keeps the coefficients unconstrained and reports the value of the coefficient on z_{t-1} and then $E_t z_{t+1}$. Optimal Weights ERME uses the variance-covariance matrix from the projection step as weights in the second step. Equally weighted ERME uses the identity matrix instead. P-value of the J-Statistic is an overall specification test. Hence rejection of the null (i.e., p-value < 0.05, say), indicates model misspecification.