Fully Modified Least Squares Estimation and Inference for Systems of Cointegrating Polynomial Regressions

Wagner, Martin

IHS Working Paper 44
January 2023

Fully Modified Least Squares Estimation and Inference for Systems of Cointegrating Polynomial Regressions

Martin Wagner

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Fully Modified Least Squares Estimation and Inference for Systems of Cointegrating Polynomial Regressions

Martin Wagner *
Department of Economics
University of Klagenfurt, Austria
&
Bank of Slovenia
Ljubljana, Slovenia
&
Institute for Advanced Studies
Vienna, Austria

Abstract

We consider fully modified least squares estimation for systems of cointegrating polynomial regressions, i. e., systems of regressions that include deterministic variables, integrated processes and their powers as regressors. The errors are allowed to be correlated across equations, over time and with the regressors. Whilst, of course, fully modified OLS and GLS estimation coincide – for any regular weighting matrix – without restrictions on the parameters and with the same regressors in all equations, this equivalence breaks down, in general, in case of parameter restrictions and/or different regressors across equations. Consequently, we discuss in detail *restricted* fully modified GLS estimators and inference based upon them.

JEL Classification: C12, C13, Q20

Keywords: Fully Modified Estimation, Cointegrating Polynomial Regression, Generalized Least Squares, Hypothesis Testing

1 Introduction

We discuss fully modified least squares estimation for *systems* of cointegrating polynomial regressions (CPRs), i. e., systems of regressions that contain deterministic variables, integrated processes and their powers as regressors. The errors are allowed to be correlated across equations, over time and with the regressors. CPRs are widely-used, e. g., in the vast environmental Kuznets curve (see, e. g., Grossman and Krueger, 1993 or Wagner, 2015) and material Kuznets curve literatures (see,

*Correspondence to: Department of Economics, University of Klagenfurt, Universitätsstrasse 65–67, 9020 Klagenfurt, Austria. Email: `Martin.Wagner@aau.at`

e. g., Grabarczyk *et al.*, 2018).[1]

The paper first provides, in Section 2.2, system version results of Wagner and Hong (2016) for fully modified OLS (FM-OLS) estimation, as originally introduced for linear cointegrating regressions by Phillips and Hansen (1990) and studied in detail for single equation CPRs in Wagner and Hong (2016).[2] The second contribution of the paper, in Section 2.3, is a detailed consideration of fully modified *generalized least squares* (FM-GLS) estimation under parameter restrictions. It is a well-known algebraic fact – first considered in Zellner (1962) for a system of seemingly unrelated regressions (SUR) in a "classical" setting – that without parameter restrictions and identical regressors in all equations, (FM-)OLS and (FM-)GLS estimation coincide for any regular weighting matrix. In case of restrictions and/or different regressors across equations GLS becomes relevant.[3] We also consider Wald-type hypothesis tests based on the restricted FM-GLS estimator and close with (i) a discussion concerning equivalence of FM-OLS and FM-GLS under restrictions and (ii) discussing the links to the seemingly unrelated cointegrating polynomial regression analysis developed in Wagner *et al.* (2020).[4]

2 Theory

2.1 Setting and Assumptions

We start with considering *unrestricted* systems of cointegrating polynomial regressions (SCPRs) where all equations include the same set of regressors:

$$\begin{aligned} y_t &= \Theta_D D_t + \Theta_X X_t + u_t, \quad t = 1, \ldots, T \\ &= \Theta Z_t + u_t \\ x_t &= x_{t-1} + v_t, \end{aligned} \tag{1}$$

[1]Other applications include the intensity-of-use literature (see, e. g., Labson and Crompton, 1993) or the exchange rate target zone literature (see, e. g., Svensson, 1992).

[2]For brevity, the focus in this paper is on FM-OLS estimation and Wald-type hypothesis testing only. Both, specification and (non-)cointegration tests are not considered here. The specification tests considered in Wagner and Hong (2016, Propositions 3 and 4) can be extended rather straightforwardly to the present setting. A system extension of (non-)cointegration tests is not feasible in general, see also Wagner (2022).

[3]From the perspective of a "big encompassing model" with all variables considered as regressors in all equations, considering different regressors in different equations simply amounts to imposing exclusion (or zero) restrictions. From this perspective, the issue of OLS and GLS equivalence boils down to restrictions being in place or not.

[4]The reference to the SUR literature is mainly due to the fact that, to the best of my knowledge, restricted fully modified estimation has not been "written down in full detail" in the single equation or systems literature that focus on OLS-type estimation. It could very well be, however, that I have missed such contributions.

with $y_t := (y_{1t}, \ldots, y_{nt})'$, $Z_t := (D_t', X_t')'$ with $D_t := (D_{1t}, \ldots, D_{qt})'$, $x_t := (x_{1t}, \ldots, x_{mt})'$, $X_t := (X_{1t}', \ldots, X_{mt}')'$, $X_{it} := (x_{it}, \ldots, x_{it}^{p_i}{}')'$ for $i = 1, \ldots, m$ and integers $p_i \geq 1$. Denoting with $p := \sum_{i=1}^{m} p_i$, the parameter matrix $\Theta = [\Theta_D, \Theta_X] \in \mathbb{R}^{n\times(q+p)}$.

For the $(n+m)$-dimensional error process $\{\xi_t\}_{t\in\mathbb{Z}} := \{[u_t', v_t']'\}_{t\in\mathbb{Z}}$ we assume that $\xi_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$ with $\sum_{j=0}^{\infty} j||C_j|| < \infty$, $\det(C(1)) \neq 0$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ a strictly stationary ergodic martingale difference sequence with natural filtration $\mathcal{F}_t = \sigma\left(\{\varepsilon_s\}_{-\infty}^{t}\right)$ and (conditional) covariance matrix $\Sigma := \mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) > 0$. Furthermore, we assume that $\sup_{t\geq 1} \mathbb{E}\left(||\varepsilon_t||^r|\mathcal{F}_{t-1}\right) < \infty$ a.s. for some $r > 4$.

For a discussion concerning these or similar sets of sufficient assumptions for the required functional limit theory see, e. g., Wagner and Hong (2016). Nonsingularity of $C(1)$ excludes cointegration in x_t, which in a linear context amounts to stating that the *triangular* system representation is based on a correctly specified dimension of the cointegrating space and this assumption is also routinely employed in, e. g., the CPR context.

The assumptions are sufficient for an invariance principle to hold for $\{\xi_t\}_{t\in\mathbb{Z}}$, i. e.,:

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \xi_t \Rightarrow B(r) = \begin{pmatrix} B_u(r) \\ B_v(r) \end{pmatrix} = \Omega^{1/2} W(r) \tag{2}$$

for $0 \leq r \leq 1$, with (by assumption) positive definite long-run covariance matrix:

$$\Omega = \begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix} := \sum_{j=-\infty}^{\infty} \mathbb{E}(\xi_{t-j}\xi_t'), \tag{3}$$

partitioned according to the partitioning of ξ_t. For later usage, we also define the half long-run covariance $\Delta := \sum_{j=0}^{\infty} \mathbb{E}(\xi_{t-j}\xi_t')$, partitioned analogously to Ω.

2.2 Estimation and Inference

Combining all observations for $t = 1, \ldots, T$ leads to:

$$Y = \Theta Z + U \tag{4}$$

$$Y' = Z'\Theta' + U', \tag{5}$$

with $Y = (y_1, \ldots, y_T)$, $Z = (Z_1, \ldots, Z_T)$ and $U = (u_1, \ldots, u_T)$. As already mentioned above, it is known since Zellner (1962) that for systems of equations (that are linear in parameters) with identical regressors in all equations, OLS estimation coincides (algebraically) with GLS estimation for any (regular) weighting matrix.[5] Consequently, without parameter restrictions and identical regressors it suffices to consider the system version of the single equation *fully modified* OLS estimator developed in Wagner and Hong (2016). For later use (when restrictions are imposed), it is convenient to also consider the vectorized versions of (4) and (5):

$$\text{vec}(Y) = (Z' \otimes I_n)\text{vec}(\Theta) + \text{vec}(U) \tag{6}$$

$$\text{vec}(Y') = (I_n \otimes Z')\text{vec}(\Theta') + \text{vec}(U'). \tag{7}$$

Equations (6) and (7) are referred to as vectorized by *observation* and *equation*, respectively. To discuss OLS and FM-OLS estimation, several (sequences of) scaling matrices need to be defined. For the deterministic regressors, we assume that there exists a scaling matrix $G_D = G_D(T)$ such that $T^{1/2}G_D D_{\lfloor rT \rfloor} \Rightarrow D(r)$ for $0 \leq r \leq 1$, with $\int_0^1 D(r)D(r)'dr$ positive definite.[6] The scaling matrix corresponding to the regressor vector X_t is given by $G_X = G_X(T) := \text{diag}(G_{X_1}, \ldots, G_{X_m})$, with $G_{X_j} := \text{diag}(T^{-1}, \ldots, T^{-\frac{p_j+1}{2}})$ for $j = 1, \ldots, m$. These matrices are combined in $G := \text{diag}(G_D, G_{X_1}, \ldots, G_{X_m})$. By definition, $T^{1/2}GZ_{\lfloor rT \rfloor} \Rightarrow J(r) := (D(r)',$
$\mathbf{B}_{v_1}(r)', \ldots, \mathbf{B}_{v_m}(r)')'$, with $\mathbf{B}_{v_j}(r) := (B_{v_j}(r), \ldots, B_{v_j}(r)^{p_j})'$, $0 \leq r \leq 1$, for $j = 1, \ldots, m$.

Clearly, as in the single equation case, the OLS estimator is also consistent in the system case with a limiting distribution that is, in general, contaminated by so-called second-order bias terms. This leads to the *fully modified* OLS (FM-OLS) estimator, which necessitates the definition of a few more quantities: First, define $y_t^+ := y_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\Delta x_t$ and $Y^+ := (y_1^+, \ldots, y_T^+)$. Note that Ω and Δ – and a fortiori their sub-blocks – can be consistently estimated based on $[\hat{u}_t', \Delta x_t']'$, with $\hat{u}_t$ denoting the OLS residuals of estimating (1). Consistency of OLS in conjunction with a usual set of assumptions on kernels and bandwidths (compare, e. g., Jansson, 2002) directly leads to consistency of long-run covariance estimation. Throughout the paper, we thus assume that long-run covariance estimation is consistent without further mentioning. Second, define an additive bias correction term:

$$\hat{M}^+ := \hat{M}_u - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{M}_v, \tag{8}$$

with:

$$\hat{M}_u := \begin{pmatrix} 0 & \ldots & 0 & \hat{\Delta}_{u_1v_1}T & 2\hat{\Delta}_{u_1v_1}\sum_{t=1}^T x_{1t} & \ldots & p_m\hat{\Delta}_{u_1v_m}\sum_{t=1}^T x_{mt}^{p_m-1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & \hat{\Delta}_{u_nv_1}T & 2\hat{\Delta}_{u_nv_1}\sum_{t=1}^T x_{1t} & \ldots & p_m\hat{\Delta}_{u_nv_m}\sum_{t=1}^T x_{mt}^{p_m-1} \end{pmatrix}$$

[5] Note that for simplicity we use the terminology GLS for any variant of weighted least squares and not only – as in a classical setting – when weighting takes place with the inverse of the covariance matrix of the errors.

[6] The standard example is $D_t = (1, t, \ldots, t^{q-1})'$ with $G_D = \text{diag}(T^{-1/2}, \ldots, T^{-(q-1/2)})$ and $T^{1/2}G_D D_{\lfloor rT \rfloor} \Rightarrow D(r) = (1, r, \ldots, r^{q-1})'$.

and $\hat{M}_v$ defined analogously to $\hat{M}_u$, with $\hat{\Delta}_{u_i v_j}$, $i = 1, \ldots, n$ replaced by $\hat{\Delta}_{v_i v_j}$, $i = 1, \ldots, m$.

Proposition 1 *Let the data be generated by* (1) *with the stated assumptions in place. The FM-OLS estimator of* Θ*, defined as:*

$$\hat{\Theta}^+ := (Y^+ Z' - \hat{M}^+)(ZZ')^{-1}, \tag{9}$$

with $\hat{M}^+$ *as given in* (8) *is consistent and its limiting distribution is given by:*[7]

$$\left(\hat{\Theta}^+ - \Theta\right) G^{-1} \Rightarrow \int_0^1 dB_{u\cdot v}(r) J(r)' \left(\int_0^1 J(r)J(r)'dr\right)^{-1}, \tag{10}$$

with $B_{u\cdot v}(r) := B_u(r) - \Omega_{uv}\Omega_{vv}^{-1} B_v(r)$.

The limiting distribution given in (10) is a zero-mean Gaussian mixture distribution since $\mathbf{B}_v(r)$ and $B_{u\cdot v}(r)$ are independent by construction. This is the basis for asymptotic standard chi-squared or normal inference. For hypothesis testing, it is convenient to consider vectorized quantities, in particular vectorized by equation. Thus, consider $\theta := \text{vec}(\Theta')$, for which (9) and (10) directly imply:

$$\begin{aligned} \hat{\theta}^+ &= \text{vec}\left((ZZ')^{-1}(ZY^{+\prime} - \hat{M}^{+\prime})\right) \\ &= \left(I_n \otimes (ZZ')^{-1}\right) \text{vec}\left(ZY^{+\prime} - \hat{M}^{+\prime}\right) \\ &= \left(I_n \otimes (ZZ')^{-1}\right) \left((I_n \otimes Z)\text{vec}(Y^{+\prime}) - \text{vec}(\hat{M}^{+\prime})\right) \end{aligned} \tag{11}$$

and:

$$\left(I_n \otimes G^{-1}\right)\left(\hat{\theta}^+ - \theta\right) \Rightarrow \text{vec}\left(\left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r) dB_{u\cdot v}(r)'\right). \tag{12}$$

The limiting distribution given in (12) is – conditional upon $J(r)$ – distributed as $\text{N}\left(0, \Omega_{u\cdot v} \otimes (\int_0^1 J(r)J(r)'dr)^{-1}\right)$. Because $\hat{\Theta}^+$, or equivalently in vectorized form $\hat{\theta}^+$, contain elements that converge at different rates, obtaining results for Wald-type test statistics requires a condition on the constraint matrix R (in case of linear hypotheses) that is not required when all estimated coefficients converge at the same rate (for more details see, e. g., Vogelsang and Wagner,

[7]This might be an unnecessary clarification since this is standard integration notation in mathematics (but less often seen in econometrics texts): The (i, j)-element of $\int_0^1 dB_{u\cdot v}(r)J(r)'$ is equal to $\int_0^1 J_j(r)dB_{u\cdot v,i}(r)$.

2014, Wagner and Hong, 2016 or Wagner *et al.*, 2020).[8] A sufficient condition on R is to assume that there exists a scaling matrix $G_R = G_R(T) \in \mathbb{R}^{s \times s}$ such that:

$$\lim_{T\to\infty} G_R^{-1} R(I_n \otimes G) = R^*, \tag{13}$$

with $R^* \in \mathbb{R}^{s\times n(q+p)}$ of full row rank s.

Proposition 2 *Let the data be generated by* (1) *with the stated assumptions in place. Consider s linearly independent linear restrictions collected in:*

$$H_0 : Rvec(\Theta') = R\theta = r, \tag{14}$$

with $R \in \mathbb{R}^{s\times n(q+p)}$ of full row rank s, $r \in \mathbb{R}^s$ and suppose there exists a scaling matrix G_R such that condition (13) *holds. Then it holds that the Wald-type statistic:*

$$T_W := (R\hat{\theta}^+ - r)' \left(R \left(\hat{\Omega}_{u\cdot v} \otimes \left(ZZ'\right)^{-1} \right) R' \right)^{-1} (R\hat{\theta}^+ - r) \tag{15}$$

is asymptotically chi-squared distributed with s degrees of freedom under the null hypothesis. In case $s = 1$, a t-type statistic:

$$T_t := \frac{R\hat{\theta}^+ - r}{\sqrt{R\left(\hat{\Omega}_{u\cdot v} \otimes (ZZ')^{-1}\right) R'}} \tag{16}$$

is asymptotically standard normally distributed under the null hypothesis.

2.3 Estimation and Inference under Restrictions

The cointegrating regression literature does not focus on restricted estimation. Consequently, since without parameter restrictions and with identical regressors OLS and GLS coincide, fully modified GLS estimation is typically not discussed. An exception is the seemingly unrelated regression cointegration literature, see, e. g., Moon (1999), Moon and Perron (2005), Park and Ogaki (1991) or Wagner *et al.* (2020). In SUR systems with potentially altogether different regressors across equations, GLS estimation becomes relevant. Using the terminology of the SUR cointegration literature, choices considered as weighting matrices include $\hat{W} = I_n$, i. e., OLS estimation, $\hat{W} = \hat{\Omega}_{uu}^{-1}$, refereed to as modified SUR estimator, or $\hat{W} = \hat{\Omega}_{u\cdot v}^{-1}$, referred to as fully modified SUR estimator by, e. g., Park and Ogaki (1991).[9]

[8] Early discussions of the matter are already contained in, e. g., Park and Phillips (1988, 1989) or Sims *et al.* (1990). Considering only linear restrictions allows to obtain closed form solutions.

[9] To be precise, the weighting matrices are given by $I_n \otimes I_T = I_{nT}$, $\hat{\Omega}_{uu}^{-1} \otimes I_T$ and $\hat{\Omega}_{u\cdot v}^{-1} \otimes I_T$ for the three considered cases.

To obtain a closed-form solution for the restricted estimator, we consider only linear restrictions on the parameter vector θ, i. e.,:

$$\theta = H\gamma + h, \tag{17}$$

with $H \in \mathbb{R}^{n(q+p)\times d}$ of full column rank d and $h \in \mathbb{R}^{n(q+p)}$.[10] Analogously to the discussion above, we need to posit a condition on the constraint matrix to arrive at a zero mean Gaussian mixture distribution. More specifically, we assume that there exists a scaling matrix $G_H = G_H(T) \in \mathbb{R}^{d\times d}$ such that:

$$\lim_{T\to\infty} (I_n \otimes G^{-1})HG_H = H^*, \tag{18}$$

with $H^* \in \mathbb{R}^{n(q+p)\times d}$ of full column rank d.

Proposition 3 *Let the data be generated by equation* (1) *with the stated assumptions in place and θ fulfilling the (explicit) restrictions posited in* (17). *Suppose that there exists a matrix G_H such that condition* (18) *holds. The* restricted fully modified generalized least squares (FM-GLS) estimator $\tilde{\theta}_R^+$ *of θ with (symmetric) weighting matrix $\hat{W}$ is defined as:*

$$\tilde{\theta}_R^+ := H\tilde{\gamma}^+ - h, \tag{19}$$

with:

$$\begin{aligned}\tilde{\gamma}^+ \;\; &:= \;\; \left(H'(\hat{W}\otimes ZZ')H\right)^{-1} \times \\ &\quad \left(H'\left(vec\left((ZY^{+\prime} - \hat{M}^{+\prime})\hat{W}\right) - (\hat{W}\otimes ZZ')h\right)\right).\end{aligned} \tag{20}$$

For $T\to\infty$ and $\hat{W}\to W>0$, it holds that:

$$\begin{aligned}G_H^{-1}\left(\tilde{\gamma}^+ - \gamma\right) \;\; &\Rightarrow \;\; \left(H^{*\prime}\left(W\otimes\int_0^1 J(r)J(r)'dr\right)H^*\right)^{-1} \times \\ &\quad \left(H^{*\prime} vec\left(\int_0^1 J(r)dB_{u\cdot v}(r)'W\right)\right).\end{aligned} \tag{21}$$

The limiting distribution of $\tilde{\gamma}^+$ given in (21) *is – conditional upon $J(r)$ – distributed as $N\left(0, \Sigma_{\tilde{\gamma}^+\tilde{\gamma}^+}\right)$, with:*

$$\begin{aligned}\Sigma_{\tilde{\gamma}^+\tilde{\gamma}^+} \;\; &:= \;\; A^{-1}BA^{-1} \\ A \;\; &:= \;\; H^{*\prime}\left(W\otimes\int_0^1 J(r)J(r)'dr\right)H^* \\ B \;\; &:= \;\; H^{*\prime}\left(W\Omega_{u\cdot v}W\otimes\int_0^1 J(r)J(r)'dr\right)H^*.\end{aligned} \tag{22}$$

[10] The two formulations of restrictions, either *implicit* via $R\theta = r$ or *explicit* via $\theta = H\gamma + h$ are, as is well-known, closely related. Starting from the explicit formulation, denote with $H_\perp \in \mathbb{R}^{n(q+p)\times n(q+p)-d}$ a matrix of full column rank that fulfills $H_\perp' H = 0$. Then, $R = H_\perp'$ and $r = H_\perp' h$ provide an implicit (re)formulation of the restrictions.

A consistent estimator of $\Sigma_{\tilde{\gamma}^+\tilde{\gamma}^+}$ is immediately available and asymptotically chi-squared or standard normally distributed test statistics follow, under conditions (18) and (24), similarly to Proposition 2.[11]

Proposition 4 *Let the data be generated by* (1) *with the assumptions discussed in place and with the true parameter vector* $\theta = H\gamma + h$*, with condition* (18) *in place, fulfilling* k *linearly independent linear restrictions, i. e.,:*

$$R_\gamma \gamma = r_\gamma, \tag{23}$$

with $R_\gamma \in \mathbb{R}^{k\times d}$ *of full column rank* k *and* $r_\gamma \in \mathbb{R}^k$*. Furthermore, assume that there exists a scaling matrix* $G_\gamma \in \mathbb{R}^{k\times k}$ *such that:*

$$\lim_{T\to\infty} G_\gamma^{-1} R_\gamma G_H \to R_\gamma^* \tag{24}$$

with $R_\gamma^* \in \mathbb{R}^{k\times d}$ *of full row rank* k *and that* $\hat{W} \to W > 0$*, then the Wald-type statistic:*

$$T_W := \left(R_\gamma \tilde{\gamma}^+ - r_\gamma\right)' \left(R_\gamma \hat{A}^{-1} \hat{B} \hat{A}^{-1} R_\gamma'\right)^{-1} \left(R_\gamma \tilde{\gamma}^+ - r_\gamma\right), \tag{25}$$

with:

$$\begin{aligned} \hat{A} &:= H'\left(\hat{W} \otimes ZZ'\right) H & (26)\\ \hat{B} &:= H'\left(\hat{W}\hat{\Omega}_{u\cdot v}\hat{W} \otimes ZZ'\right) H, & (27) \end{aligned}$$

is asymptotically chi-squared distributed with k *degrees of freedom under the null hypothesis. In case* $k = 1$ *a* t*-type statistic that is asymptotically standard normally distributed under the null hypothesis analogously to Proposition 2.*

In case $W = \Omega_{u\cdot v}^{-1}$, the asymptotic covariance matrix of $\tilde{\gamma}^+$ simplifies to:[12]

$$\Sigma_{\tilde{\gamma}^+\tilde{\gamma}^+} = \left(H^{*\prime}\left(\Omega_{u\cdot v}^{-1} \otimes \int_0^1 J(r)J(r)'dr\right) H^*\right)^{-1}. \tag{28}$$

This also implies, of course, corresponding "simplifications" of the test statistic given in Proposition 4, i. e., the test statistic defined in (25) simplifies to:

$$T_W = \left(R_\gamma \tilde{\gamma}^+ - r_\gamma\right)' \left(R_\gamma \left(H'(\hat{\Omega}_{u\cdot v}^{-1} \otimes ZZ')H\right)^{-1} R_\gamma'\right)^{-1} \left(R_\gamma \tilde{\gamma}^+ - r_\gamma\right). \tag{29}$$

[11] Note for completeness that the limiting distribution of $\tilde{\theta}_R^+$ is by definition singular.

[12] Denoting the limiting covariance matrix of $\hat{\theta}^+$ considered in Proposition 1 with $\Sigma_{\hat{\theta}^+\hat{\theta}^+}$, it follows in case $W = \Omega_{u\cdot v}^{-1}$ that $\Sigma_{\tilde{\gamma}^+\tilde{\gamma}^+} = (H^{*\prime}\Sigma_{\hat{\theta}^+\hat{\theta}^+}^{-1} H^*)^{-1}$.

Remark 1 *In the unrestricted case, i. e., when all equations contain the same set of regressors and without parameter restrictions, GLS-type estimation coincides with OLS-type estimation for any regular weight matrix. This is seen immediately by setting $H = I_{n(q+p)}$ and $h = 0$ in the expressions for the restricted estimator in Proposition 3. It is also easy to see that this equivalence extends to restricted estimation that corresponds to, e. g., eliminating the same set of regressors from all equations. In this case, it holds that $H = I_n \otimes \mathcal{H}$, with $\mathcal{H}$ a selection matrix of dimensions corresponding to the number of regressors/parameters excluded. This simplifies $\tilde{\gamma}^+$ as defined in* (20) *to:*[13]

$$\tilde{\gamma}^+ = \left(\hat{W} \otimes \mathcal{H}ZZ'\mathcal{H}'\right)^{-1} \left(vec(\mathcal{H}(ZY^{+\prime} - \hat{M}^{+\prime})\hat{W}) - (\hat{W} \otimes \mathcal{H}ZZ')h\right), \tag{30}$$

with asymptotic covariance matrix $\Sigma_{\tilde{\gamma}^+\tilde{\gamma}^+} = \Omega_{u\cdot v} \otimes \left(\mathcal{H}' \int_0^1 J(r)J(r)'dr\mathcal{H}\right)^{-1}$, which coincides exactly with the limiting distribution of the unrestricted FM-OLS estimator in the smaller model.

A similar result also prevails more generally whenever $H = I_n \otimes \mathcal{H}$, even when $\mathcal{H}$ is not a selection matrix, as long as a condition like (18) *holds with a limiting full rank matrix of the form $H^* = I_n \otimes \mathcal{H}^*$, with $\mathcal{H}^*$ not necessarily equal to $\mathcal{H}$ in general. This highlights that OLS- and GLS-type estimation coincide* also *under general linear restrictions, as long as identical restrictions that fulfill* (18) *are put on the parameters in all equations.*

Remark 2 *The results of this paper are very closely related to the seemingly unrelated CPR results of Wagner* et al. *(2020). This is most easily seen by considering, using the notation of Wagner* et al. *(2020), identical regressors in all equations, i. e., $Z_{i,t} = Z_t$ for $i = 1, \ldots, N$: Then, the OLS estimator of θ, "organized" in the same way in both papers, as defined in equation (10) of Wagner* et al. *(2020) – of course – coincides with the OLS-estimator considered (only implicitly) in this paper. Furthermore, the FM-SOLS estimator defined in equation (13) of Wagner* et al. *(2020) coincides, in the unrestricted identical regressors case, with the FM-OLS estimator defined in Proposition 1 of this paper. The FM-SUR estimator defined in equation (14) in Wagner* et al. *(2020) simplifies to FM-OLS as well in the identical regressors case, for the same (algebraic) reason as discussed for the system case in this paper. Note that Wagner* et al. *(2020) only consider a special type of parameter restrictions that they refer to as* partial pooling, *corresponding to a situation where a subset of the parameters corresponding to "similar" variables (e. g., the logarithm of GDP squared) are restricted to be identical over a subset of equations (e. g., countries). This kind of restrictions,*

[13] In this case, the matrix G_H in (18) is given by $I_n \otimes G^s$, with G^s having as its diagonal elements the elements of G corresponding to the non-excluded elements of θ. This, in turn, implies that $H = I_n \otimes \mathcal{H} = H^*$.

of course, fulfill our condition (18).

3 Summary and Conclusions

We contribute to the fully modified estimation literature in three related aspects. First, we provide the system version results of the FM-OLS estimator, as well as inference based upon it, for systems of CPRs with identical regressors in all equations. This extends the Wagner and Hong (2016) results from the single equation to the system of equations case. For this setting it follows analogously to Zellner (1962) that OLS and GLS estimation coincide for any weighting matrix. This equivalence breaks down in general in case of parameter restrictions and/or different regressors across equations. Therefore, we secondly consider in detail *restricted* FM-GLS estimators and inference based upon them. The FM-GLS estimators in general exhibit sandwich-type limiting covariance matrices unless $W = \Omega_{u \cdot v}^{-1}$. Third, we discuss conditions for asymptotic equivalence of FM-OLS and FM-GLS under restrictions and we also relate the results of this paper to the seemingly unrelated CPR analysis developed in Wagner *et al.* (2020).

References

Grabarczyk, P., M. Wagner, M. Frondel and S. Sommer (2018). A Cointegrating Polynomial Regression Analysis of the Material Kuznets Curve Hypothesis. *Resources Policy* **57**, 236–245.

Grossman, G.M. and A.B. Krueger (1993). Environmental Impacts of a North American Free Trade Agreement. In Garber, P. (Ed.) *The Mexico-US Free Trade Agreement*, 13–56, MIT Press, Cambridge.

Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory* **18**, 1449–1459.

Labson, B.S. and P.L. Crompton (1993). Common Trends in Economic Activity and Metals Demand: Cointegration and the Intensity of Use Debate. *Journal of Environmental Economics and Management* **25**, 147–161.

Moon, H.R. (1999). A Note on Fully-Modified Estimation of Seemingly Unrelated Regression Models with Integrated Regressors. *Economics Letters* **65**, 25–31.

Moon, H.R. and B. Perron (2005). Efficient Estimation of the SUR Cointegration Regression Model and Testing for Purchasing Power Parity. *Econometric Reviews* **23**, 293–323.

Park, J.Y. and M. Ogaki (1991). Seemingly Unrelated Canonical Cointegrating Regressions. *Mimeo.*

Park, J.Y. and P.C.B. Phillips (1988). Statistical Inference in Regressions with Integrated Processes: Part I. *Econometric Theory* **4**, 468–497.

Park, J.Y. and P.C.B. Phillips (1989). Statistical Inference in Regressions with Integrated Processes: Part II. *Econometric Theory* **5**, 95–131.

Phillips, P.C.B. and B.E. Hansen (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* **57**, 99–125.

Sims, C.A., J.H. Stock and M.W. Watson (1990). Inference in Linear Time Series Models with Some Unit Roots. *Econometrica* **58**, 113–144.

Svensson, L.E.O. (1992). An Interpretation of Recent Research on Exchange Rate Target Zones. *Journal of Economic Perspectives* **6**, 119–144.

Vogelsang, T.J. and M. Wagner (2014). Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrated Regressions. *Journal of Econometrics* **178**, 741–760.

Wagner, M. (2015). The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* **30**, 948–967.

Wagner, M. (2022). Residual-Based Cointegration and Non-Cointegration Tests for Cointegrating Polynomial Regressions. Forthcoming in *Empirical Economics*, https://doi.org/10.1007/s00181-022-02332-3.

Wagner, M. and S.H. Hong (2016). Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference. *Econometric Theory* **32**, 1289–1315.

Wagner, M., P. Grabarczyk and S.H. Hong (2020). Fully Modified OLS Estimation and Inference for Seemingly Unrelated Cointegrating Polynomial Regressions and the Environmental Kuznets Curve for Carbon Dioxide Emissions. *Journal of Econometrics* **214**, 216–255.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* **57**, 348–368.

Appendix: Proofs

Proof of Proposition 1: The result presents the system version of the FM-OLS estimator of θ and its asymptotic properties derived for the single equation case with $n = 1$ in Wagner and Hong (2016, Proposition 1) and follows from combining the individual equation results.

Proof of Proposition 2: Under the null hypothesis and condition (13) it holds that:

$$G_R^{-1}(R\hat{\theta}^+ - r) = G_R^{-1}R(I_n \otimes G)(I_n \otimes G^{-1})(\hat{\theta}^+ - \theta) \Rightarrow R^*\mathcal{Z}, \tag{31}$$

with $\mathcal{Z}$ denoting the random variable (limiting distribution) given in (12). Conditional upon $J(r)$, $R^*\mathcal{Z}$ is (under the null hypothesis) distributed as $\mathrm{N}\left(0, R^*(\Omega_{u\cdot v} \otimes (\int_0^1 J(r)J(r)'dr)^{-1})R^{*\prime}\right)$. Under condition (13) it furthermore holds that:

$$G_R^{-1}R\left(\hat{\Omega}_{u\cdot v} \otimes (ZZ')^{-1}\right)R'G_R^{-1} \Rightarrow R^*\left(\Omega_{u\cdot v} \otimes \left(\int_0^1 J(r)J(r)'dr\right)^{-1}\right)R^{*\prime}. \tag{32}$$

The asymptotic chi-squared null distribution for T_W as defined in (15) now follows directly from combining the above two results and noting that conditional convergence to a chi-squared distribution that is (by definition) independent of $J(r)$ amounts to unconditional convergence.

Proof of Proposition 3: Given that restricted fully modified estimation is not usually considered, we look at the OLS, GLS and FM-GLS estimators for γ in turn. Starting from (7) and inserting $\theta = H\gamma + h$ leads to:

$$\mathrm{vec}(Y') - (I_n \otimes Z')h = (I_n \otimes Z')H\gamma + \mathrm{vec}(U'), \tag{33}$$

from which, as a starting point, the OLS estimator of γ immediately follows as:[14]

$$\hat{\gamma} := \left(H'(I_n \otimes ZZ')H\right)^{-1}\left(H'\left(\mathrm{vec}(ZY') - (I_n \otimes ZZ')h\right)\right). \tag{34}$$

Analogously, also the GLS estimator of γ with weighting matrix $\hat{W}$ – more precisely $\hat{W} \otimes I_T$ – follows from standard calculations, e. g., $(\hat{W} \otimes Z)\mathrm{vec}(Y') = \mathrm{vec}(ZY'\hat{W})$:

$$\tilde{\gamma} := \left(H'(\hat{W} \otimes ZZ')H\right)^{-1}\left(H'\left(\mathrm{vec}(ZY'\hat{W}) - (\hat{W} \otimes ZZ')h\right)\right). \tag{35}$$

[14]Equation (33), of course, is algebraically a system of equations with dependent variable $\mathrm{vec}(Y') - (I_n \otimes Z')h$ and regressors $(I_n \otimes Z')H$.

Given the GLS-estimator in (35), it follows from straightforward calculations that the fully modified transformations amount to replacing $\mathrm{vec}(ZY'\hat{W})$ with $\mathrm{vec}\left((ZY^{+\prime} - \hat{M}^{+\prime})\hat{W}\right)$, which leads to $\tilde{\gamma}^{+}$ as defined in (20). Under the asymptotic full column rank condition (18), the limiting distribution given in (21) follows using similar arguments and derivations as underlying Proposition 1, itself based on Wagner and Hong (2016, Proposition 1).

Proof of Proposition 4: The result follows analogously to the result for the Wald-type statistic for linear hypotheses on θ derived in Proposition 2. An additional complication is that two asymptotic full rank conditions, one related to the matrix H relating γ and θ, given in (18), and one related to the restrictions R_{γ}, given in (24), have to hold. Also, of course, $\hat{A}$ and $\hat{B}$ need to be properly scaled to converge, compare (32) in the proof of Proposition 2.