Weighted-average least squares estimation of panel data models

Giuseppe De Luca[0000−0002−1411−2543] and
Jan R. Magnus[0000−0002−3390−639X]

Abstract This paper extends the weighted-average least squares (WALS) model averaging estimator to fixed-effects and random-effects panel data models with strictly exogenous regressors. We consider both the case where the errors are independent and identically distributed, and the case with first-order autocorrelation.

Keywords: WALS, model averaging, panel data

JEL classification: C23: C51; C52; C13

Giuseppe De Luca
University of Palermo, Italy, e-mail: `giuseppe.deluca@unipa.it`

Jan R. Magnus
Vrije Universiteit Amsterdam, The Netherlands, e-mail: `jan@janmagnus.nl`

1 Introduction

In a typical statistical project, the author considers a family of possible models, then plays a game of trial and error until one of these models comes out on top according to some criterion, and then estimates the preferred model. This, of course, is a highly dubious course of action and Professor Saleh (2006) was one of the first to warn against the dour consequences of such procedures.

To understand what is wrong, we must recognize that two actions take place, namely model selection and estimation, and that we should take the noise caused by the model selection procedure (pretesting) into account when reporting estimates and standard errors of the parameters in the preferred model. This, however, is seldom done, partly because it is difficult to do so; see, e.g., Leeb and Pötscher (2005) and McCloskey (2017).

The rapidly developing field of model averaging does make a serious attempt to combine model selection and estimation. At this time, there are three main branches of model averaging: Bayesian model averaging (BMA), frequentist model averaging (FMA), and weighted-average least squares (WALS). Useful overviews of these approaches can be found in Claeskens and Hjort (2008), Moral-Benito (2015), De Luca and Magnus (2016), and Steel (2020). Our contribution is based on the WALS approach of Magnus et al. (2010), which was originally developed in the simplest possible framework, but has since been extended in many directions.

The extension studied in the current paper is to linear and static panel data models. A panel is a two-dimensional data set, where we have observations y_{it} on units (individuals, firms, states, or other units) over a period of time. The units are indexed by $i = 1, \dots, N$ and the time periods by $t = 1, \dots, T$. In practice, we may not have observations on all units over all time periods (an unbalanced panel), and the periods may not be equally spaced over time. Our approach allows for both extensions, but we will not make this explicit in the notation.

This paper extends the WALS estimator to fixed-effects and random-effects panel data models with strictly exogenous regressors. In the fixed-effects model, we show that the WALS approach satisfies the Frisch–Waugh–Lovell theorem when partialling out an arbitrary subset of focus regressors from the unrestricted model. This result implies that, when the individual effects are treated as additional (nuisance) focus parameters, the fixed-effects WALS estimator can be easily constructed by means of the usual within transformations.

The random-effects WALS estimator builds on a special case of the feasible generalized least squares (FGLS) strategy proposed by Magnus et al. (2011) to perform WALS estimation of linear models with nonspherical errors. In this special case, we first estimate the variance components from the unrestricted model, and then perform a WALS regression based on the FGLS transformations of the original data to account for the stable equicorrelation exhibited by the one-way errors of the same unit over time.

In addition to the basic setup with independent and identically distributed (i.i.d.) errors, we also use the FGLS transformations proposed by Bhargava et al. (1982) and Baltagi and Wu (1999) to analyze an extended setup of the fixed-effects and random-effects models where errors are allowed to follow a stationary AR(1) process and observations are allowed to be unequally spaced over time.

The paper is organized as follows. In Section 2 we summarize the WALS approach, and in Section 3 we extend this approach to nonspherical errors. In Section 4 we define the class of panel data models that we wish to study. In Sections 5 and 6 we obtain the WALS estimators for the case of i.i.d. errors, and in Section 7 we discuss the extended setup with AR(1) errors. In Section 8 we present an empirical application of the fixed-effects WALS estimator to analyze the impact of legalized abortion on crime using data from Donohue and Levitt (2020). Section 9 concludes. Proofs of all propositions are in the Appendix.

2 The WALS approach

The basic framework in WALS is the homoskedastic linear model

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon, \tag{1}$$

where X_1 and X_2 are nonstochastic $n\times k_1$ and $n\times k_2$ matrices such that $X = (X_1 : X_2)$ has full column-rank $k = k_1 + k_2 < n$, and ϵ is an $n \times 1$ vector of independent errors, identically distributed as $\mathcal{N}(0, \sigma^2 I_n)$, where I_n denotes the identity matrix of order n. The two sets of regressors X_1 and X_2 play a different role: X_1 contains the k_1 'focus' regressors which we want in the model on theoretical or other grounds, while X_2 contains the k_2 'auxiliary' regressors of which we are less certain. The parameters in β_1 and β_2 are called the focus and auxiliary parameters, respectively. The focus regressors are typically variables of which we wish to study the causal effects on y or whose presence in the model is well-established from previous studies, while the auxiliary regressors are typically controls that are added to avoid omitted-variable bias, or transformations of and interactions between the regressors. We assume that $k_1 \geq 0$ and $k_2 \geq 1$. The case $k_2 = 0$ is excluded because it implies that there are no auxiliary regressors and hence no model uncertainty.

Our first step is to apply the following one-to-one transformations of X_2 and β_2:

$$Z_2 = X_2\Delta_2\Psi^{-1/2}, \qquad \gamma_2 = \Psi^{1/2}\Delta_2^{-1}\beta_2, \tag{2}$$

where Δ_2 is a diagonal $k_2 \times k_2$ matrix such that the diagonal elements of the positive definite matrix $\Psi = \Delta_2 X_2' M_1 X_2 \Delta_2 / n$ are all equal to one, $\Psi^{1/2}$ is the unique square root of Ψ, and $M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$. Next, we rescale X_1 and β_1 using

$$Z_1 = X_1\Delta_1, \qquad \gamma_1 = \Delta_1^{-1}\beta_1, \tag{3}$$

where Δ_1 is a diagonal $k_1 \times k_1$ matrix such that the diagonal elements of $Z_1'Z_1/n$ are all equal to one. Since $Z_1\gamma_1 = X_1\beta_1$ and $Z_2\gamma_2 = X_2\beta_2$, we can rewrite model (1) as

$$y = Z_1\gamma_1 + Z_2\gamma_2 + \epsilon, \tag{4}$$

where $Z_2'M_1Z_2/n = I_{k_2}$. The transformations in (2) ensure that the k_2 components of the least-squares (LS) estimator of γ_2 in model (4) are independent, while the rescaling in (3) serves only to increase the numerical accuracy of the inversion and eigenvalue routines.

The WALS estimators of γ_1 and γ_2 are obtained by averaging the LS estimators $\hat{\gamma}_{1j}$ and $\hat{\gamma}_{2j}$ resulting from the 2^{k_2} models that contain all focus regressors and a subset of the auxiliary regressors in (4):

$$\tilde{\gamma}_1 = \sum_{j=1}^{2^{k_2}} w_j \hat{\gamma}_{1j}, \qquad \tilde{\gamma}_2 = \sum_{j=1}^{2^{k_2}} w_j \hat{\gamma}_{2j}, \tag{5}$$

where the w_j are nonnegative model weights that depend only on $M_1 y$ and add up to one. Given $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, the WALS estimators of β_1 and β_2 in model (1) are:

$$\tilde{\beta}_1 = \Delta_1 \tilde{\gamma}_1, \qquad \tilde{\beta}_2 = \Delta_2 \Psi^{-1/2} \tilde{\gamma}_2. \tag{6}$$

Additional insight is obtained by writing $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ as

$$\tilde{\gamma}_1 = \hat{\gamma}_{1r} - Q\tilde{\gamma}_2, \qquad \tilde{\gamma}_2 = W\hat{\gamma}_{2u}, \tag{7}$$

where $\hat{\gamma}_{1r} = (Z_1'Z_1)^{-1}Z_1'y$ is the LS estimator of γ_1 in the fully restricted model (with $\gamma_2 = 0$), $\hat{\gamma}_{2u} = Z_2'M_1y/n$ is the LS estimator of γ_2 in the unrestricted model, $Q = (Z_1'Z_1)^{-1}Z_1'Z_2$, $W = \sum_j w_j W_j$, $W_j = I_{k_2} - S_jS_j'$ is a diagonal matrix whose diagonal elements are equal to zero or one, and S_j is a $k_2 \times r_j$ selection matrix of rank $0 \leq r_j \leq k_2$ representing the r_j exclusion restrictions implied by model j, that is, $S_j' = [I_{r_j} : 0]$ or a column-permutation thereof.

The 'equivalence theorem' of Magnus and Durbin (1999) shows that the mean squared error (MSE) matrix of $\tilde{\gamma}_1$ depends only on the MSE matrix of $\tilde{\gamma}_2$, so attention can be restricted to the estimation of γ_2. The k_2 diagonal elements λ_h of W, called the 'WALS weights,' are partial sums of the model weights w_j. Specifically, we have

$$\lambda_h = \sum_{j \in \mathcal{M}_h} w_j, \tag{8}$$

where $\mathcal{M}_h$ is the subspace of 2^{k_2-1} models that contain the hth auxiliary regressor in (4). It follows that $0 \leq \lambda_h \leq 1$, and hence the k_2 components of $\tilde{\gamma}_2$ in (7) are shrinkage estimators of the components of γ_2. Like the posterior inclusion probabilities in the context of BMA, each λ_h measures the importance of including the hth auxiliary regressor in the transformed model (4).

Since $Z_2'M_1Z_2/n = I_{k_2}$, we know that the k_2 components $(\hat{\gamma}_{2u})_h$ of $\hat{\gamma}_{2u}$ are distributed independently as

$$(\hat{\gamma}_{2u})_h \sim \mathcal{N}\left(\gamma_{2h}, \sigma^2/n\right), \tag{9}$$

where γ_{2h} denotes the hth component of γ_2. Thus, if we further restrict each λ_h to depend only on $(\hat{\gamma}_{2u})_h$, then the components of $\tilde{\gamma}_2$ are also independent, and our k_2-dimensional estimation problem for γ_2 reduces to k_2 identical one-dimensional problems of the following type: given a single observation $x \sim \mathcal{N}\left(\theta, \sigma^2/n\right)$, find a 'good' shrinkage estimator of the location parameter θ. This stylized setting is known as the 'normal location problem.' Since risk properties of shrinkage estimators are little affected by the estimation of the variance parameter (see, e.g., Danilov, 2005), we assume that σ^2 is known. An equivalent representation of the normal location problem is $x^* \sim \mathcal{N}\left(\theta^*, 1\right)$, where, with σ^2 known, $x^* = \sqrt{n}x/\sigma$ is the t-ratio for testing the hypothesis $\theta = 0$, and $\theta^* = \sqrt{n}\theta/\sigma$ is the associated 'population t-ratio'.

An obvious estimator of θ in the normal location problem is x itself, which is unbiased and consistent, and is sometimes called the 'usual' estimator. Another estimator, sometimes called the 'silly' estimator, is 0 (zero). Since $\mathrm{MSE}(x) = \sigma^2/n$ and $\mathrm{MSE}(0) = \theta^2$, we prefer the silly estimator if and only if $|\theta| < \sigma/\sqrt{n}$ or, equivalently, $|\theta^*| < 1$. No estimator dominates x uniformly in the MSE sense. However, for θ sufficiently close to zero, shrinkage estimators of the form $m(x) = \lambda(x)x$ with $0 \leq \lambda(x) \leq 1$ may achieve substantial MSE improvements by trading some bias with a variance reduction. The closeness of θ to zero should be evaluated relative to the standard deviation $\sigma/\sqrt{n}$ of x. For the silly estimator, where $\lambda(x) = 0$, we obtain a lower MSE for $|\theta^*| < 1$. More generally, the region of the parameter space where $\mathrm{MSE}(m(x)) < \mathrm{MSE}(x)$ depends on the population t-ratio θ^* and the shrinkage function $\lambda(x)$.

Based on theoretical considerations regarding admissibility, bounded risk, robustness, and optimality in terms of minimax regret, we adopt a Bayesian shrinkage approach that places a proper prior π on the population t-ratio θ^*. Under quadratic loss, the Bayes estimator of θ^* is the posterior mean $m^* = \mathbb{E}(\theta^*|x^*)$ and the implied estimator of θ is $m = \sigma m^*/\sqrt{n}$. To ensure that m enjoys attractive sampling properties as a estimator of θ, we require that the prior π on θ^* satisfies the conditions:

(C1) π is symmetric around zero;
(C2) π is positive and nonincreasing on $(0, \infty)$;
(C3) π is differentiable, except possibly at 0; and
(C4) $\psi(\theta) = -\pi'(\theta)/\pi(\theta) \to \psi_0$ as $\theta \to \infty$, where $\psi_0 \geq 0$ is some finite constant.

As discussed in De Luca et al. (2024), (C1)–(C3) are mild regularity conditions, while (C4) ensures that the prior π has bounded influence on m^*. If $\psi_0 = 0$, then (C4) implies the stronger property of 'Bayesian robustness': $x^* - m^* \to 0$ as $x^* \to \infty$, which means that prior information gets essentially ignored when the data information is sufficiently strong.

Let $\Gamma(u)$ and $B(u, v)$ denote the gamma and beta functions, respectively. Examples of priors satisfying (C1)–(C4) are the class of reflected gamma-type priors

$$\pi(\theta) = \frac{bc^{(1-a)/b}}{2\Gamma((1-a)/b)}|\theta|^{-a}e^{-c|\theta|^b} \qquad (0 \leq a < 1,\ 0 < b \leq 1,\ c > 0), \tag{10}$$

the class of the reflected beta-type priors

$$\pi(\theta) = \frac{c^{1/b} b}{2\, B(1/a - 1/b, 1/b)} (1 + c|\theta|^b)^{-1/a} \qquad (0 < a < b,\ c > 0), \tag{11}$$

the horseshoe prior (Carvalho et al., 2010)

$$\pi(\theta; c) = \frac{c\sqrt{2}}{\pi^{3/2}} \int_0^\infty \frac{e^{-t}}{\theta^2 + 2c^2 t}\, dt \qquad (c > 0), \tag{12}$$

and the log prior, also known as horseshoe-like prior (Bhadra et al., 2017),

$$\pi(\theta) = \frac{1}{2\pi c} \log(1 + c^2/\theta^2) \qquad (c > 0). \tag{13}$$

As special cases of (10), we consider the Laplace ($a = 0$, $b = 1$), Weibull ($a + b = 1$), and Subbotin ($a = 0$) priors. The normal prior ($a = 0$, $b = 2$) is excluded because it violates (C4). As special cases of (11), we consider the Pareto ($b = 1$) and the Student prior with d degrees of freedom ($a = 2/(d + 1)$, $b = 2$, $c = 1/d$). Our prior densities are also required to satisfy a notion of ignorance which we call 'neutrality,' namely not knowing whether the silly estimator of θ has a lower MSE than the usual estimator, that is, whether $|\theta^*| < 1$:

(C5) π satisfies $\int_0^1 \pi(\theta)\, d\theta = \int_1^\infty \pi(\theta)\, d\theta = 1/4$.

Together with (C1), condition (C5) fixes one free prior parameter to ensure that the events $|\theta^*| < 1$ and $|\theta^*| > 1$ are equally likely a priori. The neutral Laplace and Weibull priors are obtained for $b = \log 2$, the neutral Student prior for $d = 1$ (the Cauchy distribution), and the neutral Pareto prior for $c = 2^{a/(1-a)} - 1$. Neutrality conditions for the Subbotin, horseshoe, and log priors are solved numerically. For the Subbotin, Weibull, and Pareto priors, we fix the other free prior parameter by the minimax regret criterion for m^*, where regret is defined as the difference between the MSE of m^* and its lower bound $\theta^*/(1 + \theta^*)$ (Magnus, 2002).

The Bayesian approach is used to construct our shrinkage estimators m^* and m of θ^* and θ, but the properties of these estimators are then assessed in a classical frequentist setup. De Luca et al. (2022) studied the bias and variance of m^* in repeated samples, and proposed the plug-in maximum likelihood (ML) and double-shrinkage estimators of its sampling moments. More recently, De Luca et al. (2024) derived the finite-sample distribution of $m^* - \theta^*$, and showed that the finite-sample distribution of $m^* - \theta^*$ presents sizeable departures from normality, both in terms of skewness and of excess kurtosis. In the asymptotic theory, De Luca et al. (2024) also proved uniform $\sqrt{n}$-consistency of m and obtained its asymptotic distribution. This estimator represents our estimator of the hth component γ_{2h} of γ_2 in model (4) and its sampling properties carry over, more or less straightforwardly, to the WALS estimators $\tilde{\gamma} = (\tilde{\gamma}_1', \tilde{\gamma}_2')'$ of $\gamma = (\gamma_1', \gamma_2')'$ in (5) and the WALS estimators $\tilde{\beta} = (\tilde{\beta}_1', \tilde{\beta}_2')$ of $\beta = (\beta_1', \beta_2')'$ in (6). Specifically, under (C1)–(C4), the finite-sample distribution of the WALS estimator is generally nonnormal and the choice of prior may have sizeable effects on the estimation bias.

In large samples, these estimators are uniformly $\sqrt{n}$-consistent and their asymptotic distribution is normal only under certain conditions on β_2. The issue of inference has been addressed by De Luca et al. (2023), who proposed a simulation method for obtaining re-centered and asymmetric WALS confidence and prediction intervals based on the bias-corrected posterior mean. Although uniformly consistent estimators of the sampling distribution of shrinkage-type estimators may not exist (see Lemmas 3.1 and 3.5 in Leeb and Pötscher, 2006), the Monte Carlo simulations of De Luca et al. (2023) suggest that coverage errors of WALS intervals are small.

More detailed information on the WALS approach can be found in our 2016 survey (Magnus and De Luca, 2016) and our forthcoming monograph (De Luca and Magnus, 2025).

3 Nonspherical errors

In the previous section we provided a short summary of the WALS procedure in its simplest and most basic form. But WALS can be extended in many directions, and the current paper deals with the extension to panel data. Before we discuss panel data in more detail, we need to discuss how WALS deals with the situation where the errors are not identically and independently distributed as $\mathcal{N}(0, \sigma_\epsilon^2)$, but follow a more general distribution. Thus, we write

$$y = X_1\beta_1 + X_2\beta_2 + u, \tag{14}$$

where $u \sim \mathcal{N}(0, \Omega)$ and Ω is a positive definite (hence nonsingular) matrix. In the basic setup we have $\Omega = \sigma_\epsilon^2 I_n$, and we estimate σ_ϵ^2 from the unrestricted model, whereafter we consider it to be fixed (nonrandom).

We follow the same procedure in the more general case. Hence, we assume that Ω depends on a finite and fixed number of parameters $\alpha_1, \dots, \alpha_p$, which we estimate from the unrestricted model, and then we transform

$$y^* = X_1^*\beta_1 + X_2^*\beta_2 + \epsilon, \tag{15}$$

where

$$y^* = \Omega^{-1/2}y, \quad X_1^* = \Omega^{-1/2}X_1, \quad X_2^* = \Omega^{-1/2}X_2, \quad \epsilon = \Omega^{-1/2}u. \tag{16}$$

Next, we apply the WALS theory to the new system (15), and obtain estimates of β_1 and β_2.

4 Fixed and random effects

Our statistical panel data framework is

$$y_{it} = \nu_i + x'_{it,1}\beta_1 + x'_{it,2}\beta_2 + u_{it} \qquad (i = 1, \dots, N; \quad t = 1, \dots, T), \tag{17}$$

where $(y_{it}, x'_{it,1}, x'_{it,2})$ are the observations on the outcome of interest y_{it}, the focus regressors $x_{it,1}$, and the auxiliary regressors $x_{it,2}$; and the ν_i are individual effects capturing unobserved and time-invariant heterogeneity. The total number of observations is denoted by $n = NT$.

We can write (17) in vector form by defining

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad X_{i,1} = \begin{pmatrix} x'_{i1,1} \\ \vdots \\ x'_{iT,1} \end{pmatrix}, \quad X_{i,2} = \begin{pmatrix} x'_{i1,2} \\ \vdots \\ x'_{iT,2} \end{pmatrix}, \quad u_i = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}, \tag{18}$$

and this leads to

$$y_i = \iota_T \nu_i + X_{i,1}\beta_1 + X_{i,2}\beta_2 + u_i \qquad (i = 1, \dots, N), \tag{19}$$

where ι_T denotes the $T \times 1$ vector of ones. Throughout we shall make the following assumption.

Assumption 1 The $T \times 1$ vectors $u_1, \dots, u_N$ are i.i.d. error vectors with $\mathbb{E}(u_i) = 0$ and $\text{var}(u_i) = \sigma_u^2 V_u$.

The assumption of nonrandom regressors amounts to assuming strict exogeneity of the regressors, that is, the u_{it} must be mean-independent of past, present, and future values of $x_{it} = (x'_{it,1}, x'_{it,2})'$. This condition rules out dynamic panel data models where lags of the outcome variable are treated as predetermined regressors.

Stacking further, we let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad X_1 = \begin{pmatrix} X_{1,1} \\ \vdots \\ X_{N,1} \end{pmatrix}, \quad X_2 = \begin{pmatrix} X_{1,2} \\ \vdots \\ X_{N,2} \end{pmatrix}, \tag{20}$$

and

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}, \quad \nu = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_N \end{pmatrix}, \quad D = I_N \otimes \iota_T, \tag{21}$$

and arrive at

$$y = D\nu + X_1\beta_1 + X_2\beta_2 + u. \tag{22}$$

There are two classical approaches to the estimation of (22), depending on which assumptions are made about the individual effects ν_i; see, e.g., Mundlak (1978).

In the *fixed-effects* approach, we treat the ν_i as an additional subset of (nuisance) focus parameters to be estimated jointly with the other parameters of the model—an approach motivated by the fact that unobserved time-invariant heterogeneity is often an important source of omitted-variable bias.

Proposition 1 *Given Assumption 1, the error vector u in the fixed-effects model* (22) *satisfies* $\mathbb{E}(u) = 0$ *and* $\operatorname{var}(u) = \sigma_u^2 V$*, where* $V = I_N \otimes V_u$*.*

To implement the fixed-effects model we need to regress y on $\tilde{X}_1 = (X_1 : D)$ and X_2, where we note that the matrix $\tilde{X}_1$ is of order $NT \times (N + k_1)$. Here a possible computational problem arises, caused by the fact that the matrix $\tilde{X}_1'\tilde{X}_1$ is of order $(N + k_1) \times (N + k_1)$, which will be large if N is large, making inversion of $\tilde{X}_1'\tilde{X}_1$ computationally difficult or even infeasible. We shall show how to circumvent this problem in Section 5.

In contrast, the *random-effects* approach treats the ν_i as random variables by writing (22) as

$$y = X_1\beta_1 + X_2\beta_2 + \zeta, \qquad \zeta = D\nu + u. \tag{23}$$

Proposition 2 *Assume, in addition to Assumption 1, that the ν_i are i.i.d. with* $\mathbb{E}(\nu_i) = 0$ *and* $\operatorname{var}(\nu_i) = \sigma_\nu^2$*, and that $\{\nu_i\}$ and $\{u_{it}\}$ are independent. Then, the error vector ζ in the random-effects model* (23) *satisfies* $\mathbb{E}(\zeta) = 0$ *and* $\operatorname{var}(\zeta) = \sigma_u^2 V$*, where*

$$V = I_N \otimes \left(V_u + \alpha_1\, \iota_T \iota_T'\right)$$

and $\alpha_1 = \sigma_\nu^2/\sigma_u^2$.

In the fixed-effects model, it is easy to obtain

$$V^{-1} = I_N \otimes V_u^{-1}, \qquad V^{-1/2} = I_N \otimes V_u^{-1/2}. \tag{24}$$

In the random-effects model, we obtain

$$V^{-1} = I_N \otimes \left(V_u^{-1} - \alpha_2 V_u^{-1}\iota_T\iota_T' V_u^{-1}\right), \tag{25}$$

where

$$\alpha_2 = \frac{\sigma_\nu^2}{\sigma_u^2 + \sigma_\nu^2\, \iota_T' V_u^{-1}\iota_T} = \frac{\alpha_1}{1 + \alpha_1\, \iota_T' V_u^{-1}\iota_T}, \tag{26}$$

but to obtain explicit expressions for the square root (and its inverse) of V in the random-effects model is not straightforward, except in special cases. We shall encounter one such special case in Section 6.

5 WALS estimation of the fixed-effects model with i.i.d. errors

In the case of fixed effects and i.i.d. errors, we have $V_u = I_T$ and model (22) becomes

$$y = (D : X_1) \begin{pmatrix} \nu \\ \beta_1 \end{pmatrix} + X_2\beta_2 + \epsilon, \tag{27}$$

with $\mathbb{E}(\epsilon) = 0$ and $\text{var}(\epsilon) = \sigma_\epsilon^2 I_n$. We noted above that the matrix $\tilde{X}_1 = (D : X_1)$ has $N + k_1$ columns and that the inversion of $\tilde{X}_1{}'\tilde{X}_1$ may therefore not be feasible when N is large. Fortunately, the next proposition shows that our model-averaging approach satisfies the Frisch–Waugh–Lovell theorem when partialling out an arbitrary subset of focus regressors from the unrestricted model.[1]

Proposition 3 *Consider the model*

$$y = X_0\beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon, \tag{28}$$

where the parameter vectors β_0, β_1, and β_2 have dimensions k_0, k_1, and k_2, respectively, and the matrix $X = (X_0 : X_1 : X_2)$ has full column-rank. Premultiplying (28) *by $M_0 = I_n - X_0(X_0'X_0)^{-1}X_0'$, we obtain*

$$\breve{y} = \breve{X}_1\beta_1 + \breve{X}_2\beta_2 + \breve{\epsilon}, \tag{29}$$

where

$$\breve{y} = M_0 y, \qquad \breve{X}_1 = M_0 X_1, \qquad \breve{X}_2 = M_0 X_2, \qquad \breve{\epsilon} = M_0\epsilon.$$

Then, the WALS estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ of β_1 and β_2 in the two models are the same, and the WALS estimator of β_0 is given by

$$\tilde{\beta}_0 = (X_0'X_0)^{-1}X_0'\left(y - X_1\tilde{\beta}_1 - X_2\tilde{\beta}_2\right).$$

Specializing this result to the subset of focus regressors in the block-diagonal matrix D, this means that the individual effects can be partialled out from model (22) using the familiar within-transformation

$$v_{it}^* = v_{it} - \bar{v}_i + \bar{v} \tag{30}$$

of a variable v_{it}, where $\bar{v}_i = T^{-1}\sum_t v_{it}$ and $\bar{v} = n^{-1}\sum_i\sum_t v_{it}$.

Applying this transformation to all variables in (y, X_1, X_2), we first estimate β_1 and β_2 by fitting a WALS regression of y^* on X_1^* and X_2^* with the error variance set equal to the residual sum of squares from the unrestricted model divided by the commonly-used adjusted degrees of freedom $N(T-1) - (k-1)$ (Baltagi, 2021, Section 2.2). In WALS, this adjustment has a direct effect on the model-averaging estimates of β_1 and β_2, because the unbiased estimator of σ_ϵ^2 affects the t-ratios in the unrestricted model. Moreover, since the WALS estimates depend nonlinearly on the t-ratios of the (transformed) auxiliary coefficients, such estimates must be specified

[1] This trick only works if we think of the regressors in D as focus regressors. The result does not extend to auxiliary regressors, because the auxiliary coefficients are subject to a nonlinear shrinkage step that invalidates the Frisch–Waugh–Lovell theorem.

at the outset of the estimation procedure. Given the WALS estimates $\tilde{\beta} = (\tilde{\beta}_1', \tilde{\beta}_2')'$ of $\beta = (\beta_1', \beta_2')'$, we can then estimate the ν_i by

$$\tilde{\nu}_i = \bar{y}_i - \bar{x}_{i,1}'\tilde{\beta}_1 - \bar{x}_{i,2}'\tilde{\beta}_2, \tag{31}$$

where $\bar{x}_{i,1} = T^{-1} \sum_t x_{it,1}$ and $\bar{x}_{i,2} = T^{-1} \sum_t x_{it,2}$. The estimates resulting from this procedure are identical to those obtained from the computationally more demanding WALS regression of y on $\tilde{X}_1 = (X_1, D)$ and X_2 in model (22).

6 WALS estimation of the random-effects model with i.i.d. errors

In the case of random effects and i.i.d. errors, the variance in Proposition 2 specializes to

$$V = I_N \otimes \left(I_T + \alpha_1 \iota_T \iota_T'\right) = I_N \otimes \left((T\alpha_1 + 1)J_T + (I_T - J_T)\right), \tag{32}$$

where $J_T = \iota_T \iota_T'/T$. Since J_T and $I_T - J_T$ are idempotent and $J_T(I_T - J_T) = 0$, we obtain the following result.

Proposition 4 *The matrix V, defined in* (32) *is positive definite, and*

$$V^{-1} = I_N \otimes \left(\frac{1}{T\alpha_1 + 1} J_T + (I_T - J_T)\right)$$

and

$$V^{-1/2} = I_N \otimes \left(\frac{1}{\sqrt{T\alpha_1 + 1}} J_T + (I_T - J_T)\right).$$

It follows from Proposition 4 that, for any two vectors a and b of appropriate orders,

$$\begin{aligned} V^{-1/2}(a \otimes b) &= a \otimes \left(\frac{1}{\sqrt{T\alpha_1 + 1}} J_T b + b - J_T b\right) \\ &= a \otimes \left(b - \alpha_3 \bar{b}\, \iota_T\right), \end{aligned} \tag{33}$$

since $J_T b = \bar{b}\, \iota_T$ and letting

$$\alpha_3 = 1 - \sqrt{\frac{\sigma_\epsilon^2}{T\sigma_\nu^2 + \sigma_\epsilon^2}}. \tag{34}$$

Hence, the appropriate transformation (Baltagi, 2021, Section 2.3) of a variable v_{it} is now

$$v_{it}^* = v_{it} - \alpha_3 \bar{v}_i, \tag{35}$$

where $\bar{v}_i = T^{-1} \sum_t v_{it}$.

If the panel is unbalanced in the sense that $t = 1, \dots, T_i$ for the ith unit and the T_i are not all equal, then this transformation is still valid if we replace α_3 with α_{3i}, which we obtain from (34) by replacing T with T_i.

7 The Baltagi–Wu transformation

In Sections 5 and 6 we assumed that the errors are independent and identically distributed. Let us now consider the case where V_u is not equal to the identity matrix. We shall only consider the model with stationary AR(1) errors, given by

$$u_{it} = \rho\, u_{i,t-1} + \epsilon_{it}, \tag{36}$$

where $|\rho| < 1$ and the ϵ_{it} are i.i.d. errors that follow a $\mathcal{N}(0, \sigma_\epsilon^2)$ distribution.

The variance matrix depends on how we specify the error in the first time period. The typical assumptions are

$$u_{i1} = \frac{\epsilon_{i1}}{\sqrt{1-\rho^2}} \qquad \text{or} \qquad u_{i1} = \epsilon_{i1}. \tag{37}$$

In the first (more common) case, the variance matrix is $\Omega = \sigma_\epsilon^2 (I_N \otimes V_u)$, where

$$V_u = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-3} & \rho^{T-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & \rho & 1 \end{pmatrix} \tag{38}$$

and

$$V_u^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}. \tag{39}$$

There are various ways to estimate ρ. In principle, the theory of Section 3 applies whether or not the matrix V_u is equal to the identity matrix, and it is sufficient to premultiply the system by $V^{-1/2} = I_N \otimes V_u^{-1/2}$. But in practice, it is more efficient to use a simple transformation without calculating the large matrix $V^{-1/2}$. Feasible GLS strategies for the case of AR(1) errors for both fixed effects and random effects were developed by Bhargava et al. (1982), Baltagi and Li (1991), and Baltagi and Wu (1999). In particular, the random-effects estimator of Baltagi and Wu (1999) treats exogenously unbalanced panels with unequally-spaced observations.

Based on the Baltagi–Wu transformation, we proceed as follows:

Fixed effects:

1. estimate ρ and σ_ϵ^2 in the unrestricted model;
2. use the estimated correlation $\hat{\rho}$ from step 1 to compute the Baltagi–Wu AR(1) transformations (y^*, X_1^*, X_2^*) of (y, X_1, X_2);
3. compute the within-transformations $(y^{**}, X_1^{**}, X_2^{**})$ of (y^*, X_1^*, X_2^*) by excluding the N observations corresponding to the first period of each unit;
4. estimate β_1 and β_2 by performing a WALS regression of y^{**} on X_1^{**} and X_2^{**} with error variance set equal to its unbiased estimate $\hat{\sigma}_\epsilon^2$ from step 1.

Random effects:

1. estimate ρ, σ_ϵ^2, and σ_ν^2 in the unrestricted model;
2. use the estimated correlation $\hat{\rho}$ from step 1 to compute the Baltagi–Wu AR(1) transformations (y^*, X_1^*, X_2^*) of (y, X_1, X_2);
3. use the estimated variance components $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\nu^2$ from step 1 to compute the Baltagi–Wu feasible GLS transformations $(y^{**}, X_1^{**}, X_2^{**})$ of (y^*, X_1^*, X_2^*);
4. estimate β_1 and β_2 by performing a WALS regression of y^{**} on X_1^{**} and X_2^{**}.

8 The impact of legalized abortion on crime

As an empirical illustration we revisit the abortion-crime hypothesis put forward by Donohue and Levitt (2001), which states that there exists a causal relationship between the legalization of abortion in the US during the early 1970s and the subsequent decline in violent, property, and murder crimes observed in the 1990s. Donohue and Levitt, henceforth DL, provide two main arguments for their hypothesis: a cohort-size effect and a selection effect. First, holding the number of pregnancies constant, a higher abortion rate in a given year yields a smaller youth population in the next generation. Since younger individuals tend to commit more crimes than older individuals, this decrease in the share of younger people may result in lower crime rates as the cohorts exposed to legalized abortion reach the age when crimes are typically committed. Second, legalized abortion provides women with greater control over child planning, increasing the likelihood that children are born into more favorable environments for their development. This improved selection effect may further contribute to a reduction in crime rates, even if fertility rates remain constant.

The empirical findings of DL (2001) attracted much attention, both within and outside the academic community, and led to many comments, replies, and controversies, which were reviewed in DL (2020). In this more recent study, DL also present additional empirical evidence for the causal impact of abortion on crime by extending their original state-level panel data analysis from the period 1985–1997 to the period 1985–2014. Although the new data differ from the original data in several important respects, the baseline models in the two studies are quite similar. In addition to state and year fixed effects, both studies rely on a limited set of control variables for eight time-varying and state-specific factors: log of lagged prisoners per capita, log of lagged police per capita, log of per capita income, per capita beer consumption, unemployment rate, poverty rate, generosity of the AFDC welfare program at time $t - 15$, and an indicator for the presence of concealed weapons laws.

As argued by Belloni et al. (2014a), a key concern with these parsimonious models is that state-level abortion rates in the early 1970s were not randomly assigned. Thus, failing to adequately control for factors associated with both state-level abortion and crime rates may introduce serious omitted-variable bias. Examples of such confounding factors include nonlinear aggregate trends and persistent state-to-state differences in policies, attitudes, and demographics. In the spirit of Belloni et al. (2014a,b), Hahn et al. (2018), and De Luca et al. (2022), our objective is to assess the robustness of the cross-state regression results reported by DL (2020) with respect to additional controls for suitable transformations of the available explanatory variables and their interactions with a nonlinear trend. This leads us to consider an extended specification of the fixed-effects models estimated by DL (2020) with a much richer set of control variables, thus strengthening the plausibility of the conditional independence assumption for the abortion rate. Furthermore, by using the fixed-effects WALS estimator, we can properly deal with the bias-precision trade-off in estimating the treatment effects of interest.

8.1 Preliminary analysis of the baseline specifications

Our analysis is based on a panel of 51 US states over the period 1985–2014, thus containing a total of $51 \times 30 = 1530$ state-year observations. We distinguish between four crime measures: violent crime, property crime, and murder crime based on both Uniform Crime Reports (UCR) and Vital Statistics (VS) data.[2] For each crime measure, DL (2020) estimate a fixed-effects model with stationary AR(1) errors where the dependent variable is the log of the per capita crime rate in state i at time t. The treatment variables of interest are obtained by interacting the 'effective abortion rate' (EAR) for a given state, year and crime category with binary indicators for the time periods 1985–1997 and 1998–2014.[3] Each model includes state and year fixed effects to control for time-invariant state-specific characteristics and national aggregate trends, respectively. In what follows, we focus on the specifications corresponding to columns (2), (4), (6) and (8) of DL (2020, Table 4), which also include the eight time-varying and state-specific factors described above.

In the first panel (method A) of Table 1, we present our replication in Stata of their FGLS estimates. Except for a negligible difference in the estimated coefficient of EAR '86–'97 for murder crimes based on VS data, our estimates coincide with

[2] Data are available at the website: `https://works.bepress.com/john_donohue/192/`

[3] The EAR represents a measure of the abortion rate that is relevant for each type of crime, based on the typical ages at which individuals tend to commit these crimes. For a given crime category, the indicator is defined as a weighted average of the number of abortions per 1,000 live births, with weights determined by the fractions of total arrests in 1985 across various ages (see DL, 2020, Eqn 1).

Table 1 Fixed-effects estimates of the baseline specifications for log of per capita crime rates

		Type of crime[c]			
Method[a]	Variable[b]	Violent	Property	Murder (UCR)	Murder (VS)
(A)	EAR '85–'97	−0.178 **	−0.152 **	−0.100 *	−0.117 **
		[0.022]	[0.016]	[0.040]	[0.036]
	EAR '98–'14	−0.189 **	−0.168 **	−0.152 **	−0.164 **
		[0.019]	[0.015]	[0.021]	[0.019]
	n	1517	1517	1517	1499
	k	40	40	40	40
	$\hat{\sigma}_\epsilon$	0.050	0.037	0.124	0.109
	$\hat{\rho}$	0.904	0.861	0.430	0.438
(B)	EAR '86–'97	−0.214 **	−0.163 **	−0.131 **	−0.171 **
		(0.035)	(0.027)	(0.042)	(0.038)
	EAR '98–'14	−0.223 **	−0.176 **	−0.172 **	−0.194 **
		(0.032)	(0.026)	(0.026)	(0.025)
	n	1466	1466	1466	1455
	k	39	39	39	39
	$\hat{\sigma}_\epsilon$	0.050	0.035	0.118	0.102
	$\hat{\rho}$	0.891	0.883	0.633	0.678

[a] Method (A) provides our replication in Stata of the FGLS estimates from columns (2), (4), (6) and (8) of Table 4 in DL (2020). Method (B) uses the same specifications as method (A), but the underlying estimation procedure differs in the choice of the weights and the estimation method for ρ, the use of Baltagi–Wu and Cochrane–Orcutt transformations, and the use of classical standard errors rather than robust standard errors. Details are given in the text.
[b] Model statistics: n represents the effective sample size in the second-step of the FGLS procedure, k is the number of regressors (excluding fixed effects), $\hat{\sigma}_\epsilon$ denotes the estimated standard deviation of the error term, and $\hat{\rho}$ is the estimated autocorrelation coefficient.
[c] The symbols †, $*$ and $**$ denote significance levels of 10%, 5%, and 1%, respectively. Robust standard errors are presented in squared brackets, while classical standard errors are reported in curved brackets.

those reported in DL (2020, Table 4). These fixed-effects regressions are based on a weighted two-step procedure that accounts for both state population weights and first-order serial correlation using the FGLS approach of Bhargava et al. (1982). In the first step, we compute the residuals from a weighted LS (WLS) regression with weights equal to the state population, and then estimate the correlation coefficients ρ by $\hat{\rho} = 1 - d_p/2$, where d_p is the panel data generalization of the Durbin–Watson statistics given in Bhargava et al. (1982, Eqn 4). In the second step, we perform a WLS regression with robust standard errors based on the Prais–Winsten transformations (Bhargava et al. 1982, Eqs 16a and 16b). The estimated coefficients of the two EAR variables are negative and significant at the 1% level (except the first-period effect on UCR murder, which is significant only at the 5% level). Moreover, the estimated effects for the second period are always larger than those for the first period. The empirical findings thus seem to support the abortion-crime hypothesis.

The FGLS approach employed by DL is not entirely consistent with the approach used in the development of the WALS estimator for fixed-effects models with AR(1) errors. Therefore, to obtain meaningful comparisons, it is important to assess the influence of the underlying methodological differences. This is done in the second panel (method B) of Table 1, where we present the estimates resulting from our FGLS approach.

Method (B) differs from method (A) in five respects. First, the fixed-effects algorithm described in Section 7 can be easily extended to weighted data, provided that the weights are constant within units of the panel, so that the within transformation effectively removes the fixed effects. In contrast, a dummy-variable specification of fixed effects allows for time-varying weights, but is feasible only when N (the number of states) is relatively small. The weights employed in method (B) are the average state population over the period 1985–2014. Intermediate comparisons reveal that, other things being equal, this difference has only a small effect on the estimated coefficients of the two EAR variables and their standard errors.

Second, using weighted data restricts the available methods for the first-step estimation of the autocorrelation coefficient ρ. Since the method used by DL (2020) is designed for unweighted data, we opt for another estimation method that properly accounts for the weighted data structure using a WLS regression of the first-step residuals on their lagged values.[4] The results are again similar, except in the estimates of ρ in the two models for murder crimes.

Third, our fixed-effects approach relies on the Baltagi–Wu AR(1) transformation, which can accommodate unbalanced panels with unequally spaced observations over time. For equally spaced panels, this transformation is equivalent to the panel data generalization of the Prais–Winsten transformation proposed by Bhargava et al. (1982). However, when the panel is unequally spaced, the Prais–Winsten transformation is invalid. In the current application, this issue may affect the estimates of the model for murder crime based on VS data, which are unbalanced and unequally spaced due to the presence of missing values in the outcome variable.[5]

Fourth, to ensure that the within transformation effectively removes the fixed effects, we must exclude the N observations corresponding to the first period of each unit. The differences in the sample size and the number of regressors between methods (A) and (B) are due to the use of the Cochrane–Orcutt transformation instead of the Prais–Winsten transformation. To highlight this adjustment, in method (B), we have relabeled the first-period treatment variable as EAR '86–'97.

Fifth, since method (B) will serve as the benchmark when we compare with the fixed-effects WALS estimates, we report classical standard errors. Our model averaging approach does not currently support the calculation of robust standard errors. In fact, to our knowledge, the problem of computing robust standard errors for model averaging estimators is still unexplored.

[4] Our approach is one of the two estimation methods for ρ available in the `xtregar` Stata command. The alternative method, involving a WLS regression of the residuals on their lead values, yields similar estimates of ρ.

[5] The panels for the other three types of crime are unbalanced but equally spaced, because only few values are missing in the log of lagged prisoners per capita.

When comparing the estimated coefficients of the two EAR variables, we conclude that the methodological differences between methods (A) and (B) provide additional support for the abortion-crime hypothesis: the point estimates from method (B) are consistently larger (in absolute value) than those obtained with method (A). The underlying standard errors are also relatively larger, especially in the models with sizeable first-step estimates of ρ. However, all effects are now significant at the 1% level. The central issue remains to determine whether the strong negative association between EAR and crime rates reflects a causal relationship or whether it is driven by other confounding factors.

8.2 Expanding the set of control variables

To mitigate concerns about spurious correlations, we now consider an extended specification of the models estimated by DL (2020), adding control variables for lagged values, squared terms, and interaction terms of the eight time-varying and state-specific factors already present in their baseline specifications.[6] Our set of control variables also contains interaction terms of a cubic trend with: (i) the initial level of each crime-specific EAR variable; (ii) the current levels, lagged levels, initial levels, and within-state averages of the eight time-varying and state-specific factors considered by DL; and (iii) the squares of all the aforementioned variables. After excluding some regressors due to perfect collinearity, we have a total of 323 explanatory variables: 1 constant term, 2 treatment variables, 50 state effects, 28 year effects, and 242 time-varying and state-specific controls.

Table 2 presents the fixed-effects estimates resulting from three alternative methods: method (C) uses the same fixed-effects FGLS approach as method (B) with the extended set of controls, method (D) is a pretest version of this approach based on a general-to-specific variable selection procedure, while method (E) applies the fixed-effects WALS estimator. In methods (D) and (E), the focus regressors includes the two EAR treatment variables, along with state and year fixed effects. Model selection and model averaging are performed on the remaining set of 242 auxiliary regressors. In method (C), which represents the unrestricted model, the distinction between focus and auxiliary regressors does not play a role.

Comparing the results from methods (B) and (C), we see that using a larger set of control variables leads to smaller estimates of both the error variance and the autocorrelation coefficient across all models. These differences are not entirely surprising, as method (C) uses a broader set of controls that also includes lagged levels of the original controls selected by DL and their interactions. The reduction in the first-step estimates of ρ is more pronounced in the two models for murder crime, where $\hat{\rho}$ decreases from 0.633 to 0.178 for UCR data and from 0.678 to 0.165 for VS data. For violent crimes, we find that the estimated effects of the two EAR variables

[6] We exclude the lagged generosity of the AFDC welfare program to avoid introducing additional missing values. Squared terms are limited to continuous regressors, while interaction terms are computed for both current and lagged levels of the original control variables.

Table 2 Fixed-effects estimates of the extended models for log of per capita crime rates

Method[a]	Variable[b]	Type of crime[c] Violent	Property	Murder (UCR)	Murder (VS)
(C)	EAR ’86–’97	0.001	−0.068 *	−0.146 *	−0.184 **
		(0.058)	(0.034)	(0.058)	(0.049)
	EAR ’98–’14	0.000	−0.072 *	−0.176 **	−0.208 **
		(0.056)	(0.034)	(0.048)	(0.040)
	n	1466	1466	1466	1455
	k	273	273	273	273
	$\hat{\sigma}_\epsilon$	0.046	0.032	0.105	0.089
	$\hat{\rho}$	0.762	0.727	0.178	0.165
(D)	EAR ’86–’97	0.027	−0.051 †	−0.099 *	−0.116 **
		(0.040)	(0.029)	(0.043)	(0.037)
	EAR ’98–’14	0.025	−0.056 †	−0.125 **	−0.144 **
		(0.038)	(0.029)	(0.034)	(0.028)
	n	1466	1466	1466	1455
	k_1	31	31	31	31
	k_2^*	62	116	98	90
	$\hat{\sigma}_\epsilon$	0.045	0.031	0.103	0.088
	$\hat{\rho}$	0.792	0.769	0.251	0.258
(E)	EAR ’86–’97	−0.039	−0.085 *	−0.130 *	−0.171 **
		(0.047)	(0.028)	(0.049)	(0.043)
	EAR ’98–’14	−0.042	−0.092 *	−0.162 **	−0.201 **
		(0.046)	(0.027)	(0.040)	(0.036)
	n	1466	1466	1466	1455
	k_1	31	31	31	31
	k_2	242	242	242	242
	$\hat{\sigma}_\epsilon$	0.046	0.032	0.105	0.089
	$\hat{\rho}$	0.762	0.727	0.178	0.165

[a] Method (C) employs the same fixed-effects FGLS approach as method (B). Method (D) is a pretest version of this approach based on a general-to-specific variable selection procedure. Method (E) applies the fixed-effects WALS approach with AR(1) errors and Pareto prior.
[b] Model statistics: n represents the effective sample size in the second-step of the FGLS procedure, k is the number of regressors (excluding fixed effects), k_1 is the number of focus regressors (excluding fixed effects), k_2 is the number of auxiliary regressors, k_2^* is the number of selected auxiliary regressors, $\hat{\sigma}_\epsilon$ denotes the estimated standard deviation of the error term, and $\hat{\rho}$ is the estimated autocorrelation coefficient.
[c] The symbols †, * and ** denote a significance level of 10%, 5%, and 1%, respectively. Classical standard errors are presented in curved brackets.

are positive and not significant at the conventional levels. For property crimes, the EAR coefficients decrease from −0.163 to −0.068 for the first time period and from −0.176 to −0.072 for the second time period. Moreover, each effect is significant only at the 5% level. On the other hand, for UCR and VS murder crimes, the negative effects of EAR become slightly stronger and remain significant at the 1% level.

An obvious concern with the unrestricted estimates in method (C) is the risk of overfitting, which is reflected in the large standard errors. Method (D) addresses this issue using an iterative general-to-specific variable selection procedure that retains all focus regressors and progressively eliminates the least significant auxiliary regressors if their p-value exceeds 0.05. The number of auxiliary regressors selected by this procedure ranges from a minimum of 62 controls for violent crimes to a maximum of 116 controls for property crimes. Looking at the fixed-effects FGLS estimates of the selected model for each crime measure, we see that the effects of EAR on violent crimes remain positive and not statistically significant at the conventional levels, while the effects on property and murder crimes are relatively smaller (in absolute value) than those obtained in method (C). Although method (D) produces smaller standard errors than method (C), the effects of the two EAR variables on property crimes are significant only at the 10% level and the effect of the first EAR variable on (UCR) murder crimes is significant only at the 5% level.

Model selection methods, like method (D), have several limitations. Most importantly, the properties of an estimator are often reported as if estimation had not been preceded by one or many model selection steps. However, disregarding the uncertainty due to the model selection process can lead to misleading inference (see, e.g., Leeb and Pötscher 2005). The WALS approach developed in this paper resolves some of the issues faced by other estimation methods. First, in addition to accounting for unobserved time-invariant heterogeneity and serial correlation in panel data models, the WALS approach relies on a preliminary semi-orthogonal transformation of the auxiliary regressors that greatly reduces the computational time required for model averaging. Second, it addresses the bias-precision trade-off arising in finite-sample estimation problems through a transparent Bayesian shrinkage procedure that places a prior on the population t-ratios associated with the (transformed) auxiliary coefficients. Third, the WALS estimator enjoys desirable theoretical properties such as admissibility, bounded risk, robustness, minimax regret optimality, and $\sqrt{n}$-uniform consistency. Fourth, although its finite-sample distribution is not normal, it is relatively straightforward to derive plug-in estimators of its sampling moments and simulation-based confidence intervals.

In the last panel (method E) of Table 2, we present the fixed-effects WALS estimates based on a Pareto prior and the plug-in double shrinkage estimates of their standard errors. Significance levels of the estimated coefficients are derived from simulation-based confidence intervals constructed through 5,000 Monte Carlo replications.[7] Our results suggest that the EAR coefficients are negative for all crime outcomes, but the magnitude and statistical significance of the estimates for violent and property crimes differ substantially from those obtained in method (B), which is based on the small set of controls selected by DL. In the model for violent crimes, the 90% confidence intervals for the EAR coefficients are $(-0.098, 0.089)$ for the first time period and $(-0.096, 0.085)$ for the second time period. In the model for property crimes, the EAR coefficients lie between those from methods (B) and (C), but are much closer to the unrestricted estimates from method (C). Although the WALS

[7] Using a laptop with an Intel(R) Core(TM) i7-11800H CPU/2.30 GHz processor and 32 GB of RAM, the WALS estimates of each model are computed in about 1 minute.

estimates are generally more precise than the unrestricted estimates, the estimated effects of EAR on property crimes are only significant at the 5% level. On the other hand, the estimated effects of EAR on UCR and VS murder crimes are quite close to the estimates obtained in method (B). Thus, even after conditioning on our large set of controls, the negative impact of legalized abortion on murder crimes remains strong and significant at the 1% level.

9 Concluding remarks

In this paper, we have extended the WALS approach of Magnus et al. (2010) to linear and static panel data models. We explored both fixed-effects and random-effects models, considering two alternative setups for the regression errors: one assuming i.i.d. errors and another accounting for AR(1) errors. Our fixed-effects estimation procedure exploits the fact that WALS satisfies the Frisch–Waugh–Lovell theorem when partialling out an arbitrary subset of focus regressors from the unrestricted model. Consequently, when treating the dummy variables for fixed effects as (nuisance) focus regressors that are needed to account for unobserved time-invariant heterogeneity, this subset can be partialled out from the unrestricted model using the familiar within-transformation.

The random-effects estimator and the extended setups for AR(1) errors build on special cases of the FGLS strategy for WALS estimation of linear models with nonspherical errors. Although these models are well-established in the panel data literature, the underlying WALS estimation procedure is challenging, particularly when dealing with unbalanced and unequally spaced panels.

To address these issues, we developed user-friendly Stata commands for fixed-effects and random-effects WALS estimation of the panel data models discussed in this paper. Additional details about these commands will be provided in our forthcoming monograph (De Luca and Magnus 2005), but the software is already freely available from the authors upon request.

In our empirical application, we employed the WALS estimator for fixed-effects panel data models with AR(1) errors to analyze the impact of legalized abortion on violent, property, and murder crimes, using data from Donohue and Levitt (2020). Our results provide further empirical support for Donohue and Levitt's (2001, 2020) hypothesis concerning the impact of legalized abortion on murder crimes. However, for violent and property crimes, the empirical evidence supporting this hypothesis is less definitive.

Proofs of the propositions

Proof of Proposition 1: This follows immediately from Assumption 1.

Proof of Proposition 2: It follows from our assumptions that $\mathbb{E}(\nu) = 0$, $\mathbb{E}(u) = 0$, and

$$\operatorname{var}(\nu) = \sigma_\nu^2 I_N, \qquad \operatorname{var}(u) = \sigma_u^2 I_N \otimes V_u.$$

Hence, $\mathbb{E}(\zeta) = 0$ and

$$\begin{aligned}\operatorname{var}(\zeta) &= \operatorname{var}(D\nu + u) = D\operatorname{var}(\nu)D' + \operatorname{var}(u)\\ &= \sigma_\nu^2(I_N \otimes \iota_T)(I_N \otimes \iota_T') + \sigma_u^2(I_N \otimes V_u)\\ &= \sigma_\nu^2(I_N \otimes \iota_T\iota_T') + \sigma_u^2(I_N \otimes V_u) = I_N \otimes \left(\sigma_\nu^2\iota_T\iota_T' + \sigma_u^2 V_u\right),\end{aligned}$$

thus completing the proof.

Proof of Proposition 3: Consider first model (29). Defining

$$\breve{M}_1 = I_n - \breve{P}_1, \qquad \breve{P}_1 = \breve{X}_1(\breve{X}_1'\breve{X}_1)^{-1}\breve{X}_1',$$

we have

$$\breve{P}_1 = M_0X_1(X_1'M_0X_1)^{-1}X_1'M_0$$

and

$$\breve{X}_2'\breve{M}_1\breve{X}_2 = X_2'M_0X_2 - X_2'M_0X_1(X_1'M_0X_1)^{-1}X_1'M_0X_2. \tag{40}$$

Let $\breve{\Delta}_1$ and $\breve{\Delta}_2$ be diagonal matrices with diagonal elements

$$(\breve{\Delta}_1)_{jj} = \sqrt{\frac{n}{(\breve{X}_1'\breve{X}_1)_{jj}}} \qquad (j = 1, \dots, k_1)$$

and

$$(\breve{\Delta}_2)_{hh} = \sqrt{\frac{n}{(\breve{X}_2'\breve{M}_1\breve{X}_2)_{hh}}} \qquad (h = 1, \dots, k_2),$$

respectively, implying that the diagonal elements of the matrices $\breve{\Delta}_1\breve{X}_1'\breve{X}_1\breve{\Delta}_1/n$ and $\breve{\Psi} = \breve{\Delta}_2\breve{X}_2'\breve{M}_1\breve{X}_2\breve{\Delta}_2/n$ are all equal to 1, and define

$$\breve{Z}_1 = \breve{X}_1\breve{\Delta}_1, \qquad \breve{Z}_2 = \breve{X}_2\breve{\Delta}_2\breve{\Psi}^{-1/2}, \qquad \breve{\gamma}_1 = \breve{\Delta}_1^{-1}\beta_1, \qquad \breve{\gamma}_2 = \breve{\Psi}^{1/2}\breve{\Delta}_2^{-1}\beta_2.$$

Then we obtain $\breve{Z}_1\breve{\gamma}_1 = \breve{X}_1\beta_1$ and $\breve{Z}_2\breve{\gamma}_2 = \breve{X}_2\beta_2$, which shows that model (29) can be written equivalently as

$$\breve{y} = \breve{Z}_1\breve{\gamma}_1 + \breve{Z}_2\breve{\gamma}_2 + \breve{\epsilon}, \tag{41}$$

where, by construction, $\breve{Z}_2'\breve{M}_1\breve{Z}_2/n = I_{k_2}$.

Next, consider model (28). Letting

$$X_* = (X_0 : X_1), \qquad \beta_* = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix},$$

we can write model (28) as

$$y = X_*\beta_* + X_2\beta_2 + \epsilon.$$

Now,

$$(X_*'X_*)^{-1} = \begin{pmatrix} X_0'X_0 & X_0'X_1 \\ X_1'X_0 & X_1'X_1 \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma^{00} & \Sigma^{01} \\ \Sigma^{10} & \Sigma^{11} \end{pmatrix},$$

with

$$\begin{aligned} \Sigma^{00} &= (X_0'X_0)^{-1} + (X_0'X_0)^{-1}X_0'X_1\Sigma^{11}X_1'X_0(X_0'X_0)^{-1}, \\ \Sigma^{01} &= -(X_0'X_0)^{-1}X_0'X_1\Sigma^{11}, \\ \Sigma^{11} &= (X_1'M_0X_1)^{-1}. \end{aligned}$$

Letting $P_0 = I_n - M_0 = X_0(X_0'X_0)^{-1}X_0'$ and $P_* = X_*(X_*'X_*)^{-1}X_*'$, we then obtain

$$\begin{aligned} P_* &= X_0\Sigma^{00}X_0' + X_0\Sigma^{01}X_1' + X_1\Sigma^{10}X_0' + X_1\Sigma^{11}X_1' \\ &= P_0 + P_0X_1\Sigma^{11}X_1'P_0 - P_0X_1\Sigma^{11}X_1' - X_1\Sigma^{11}X_1'P_0 + X_1\Sigma^{11}X_1' \\ &= P_0 + M_0X_1(X_1'M_0X_1)^{-1}X_1'M_0 \end{aligned}$$

and, using (40),

$$X_2'M_*X_2 = X_2'M_0X_2 - X_2'M_0X_1(X_1'M_0X_1)^{-1}X_1'M_0X_2 = \check{X}_2'\check{M}_1\check{X}_2, \qquad (42)$$

where $M_* = I_n - P_*$. This shows that the semi-orthogonal transformation of $\check{X}_2$ in model (29) has the same effect as the semi-orthogonal transformation of X_2 in model (28). In particular, letting Δ_* be the diagonal matrix

$$\Delta_* = \begin{pmatrix} \Delta_{*0} & 0 \\ 0 & \Delta_{*1} \end{pmatrix},$$

where

$$(\Delta_{*0})_{ii} = \sqrt{\frac{n}{(X_0'X_0)_{ii}}} \qquad (i = 1, \dots, k_0)$$

and

$$(\Delta_{*1})_{jj} = \sqrt{\frac{n}{(X_1'X_1)_{jj}}} \qquad (j = 1, \dots, k_1),$$

we define

$$Z_* = X_*\Delta_*, \quad Z_2 = X_2\Delta_2\Psi^{-1/2}, \quad \gamma_* = \Delta_*^{-1}\beta_*, \quad \gamma_2 = \Psi^{1/2}\Delta_2^{-1}\beta_2,$$

and obtain

$$Z_*\gamma_* = X_*\beta_*, \qquad Z_2\gamma_2 = X_2\beta_2, \qquad \frac{Z_2'M_*Z_2}{n} = \frac{\check{Z}_2'\check{M}_1\check{Z}_2}{n} = I_{k_2}.$$

Model (28) can thus be written equivalently as

$$y = Z_*\gamma_* + Z_2\gamma_2 + \epsilon. \tag{43}$$

Consider next the WALS estimators of $\breve{\gamma}_2$ and γ_2 in models (41) and (43). The underlying LS estimators are the same, because

$$\begin{aligned} Z_2'M_*y &= Z_2'\left(M_0 - M_0X_1(X_1'M_0X_1)^{-1}X_1'M_0\right)y \\ &= Z_2'M_0\left(I_n - M_0X_1(X_1'M_0X_1)^{-1}X_1'M_0\right)M_0y = \breve{Z}_2'\breve{M}_1\breve{y}, \end{aligned}$$

which ensures that, after taking into account the degrees of freedom in estimating σ_ϵ^2, the k_2-vectors of t-ratios in the two models are the same and so are the underlying WALS estimators of γ_2. The same is true for the WALS estimators of β_2 in models (28) and (29) because the transformations in (41) and (43) are based on the same Δ_2 and Ψ, in view of (42). Regarding the WALS estimators of the focus parameters, we note that

$$\begin{aligned} \breve{Q} &= (\breve{Z}_1'\breve{Z}_1)^{-1}\breve{Z}_1'\breve{Z}_2 = \breve{\Delta}_1^{-1}(X_1'M_0X_1)^{-1}X_1'M_0X_2\breve{\Delta}_2\breve{\Psi}^{-1/2}, \\ \hat{\gamma}_{1r} &= (\breve{Z}_1'\breve{Z}_1)^{-1}\breve{Z}_1'\breve{y} = \breve{\Delta}_1^{-1}(X_1'M_0X_1)^{-1}X_1'M_0y, \end{aligned}$$

leading to the WALS estimator of β_1 in model (29):

$$\tilde{\beta}_1 = \breve{\Delta}_1(\hat{\gamma}_{1r} - \breve{Q}\tilde{\gamma}_2) = (X_1'M_0X_1)^{-1}X_1'M_0(y - X_2\tilde{\beta}_2).$$

In model (28) we have

$$(Z_*'Z_*)^{-1}Z_*' = \Delta_*^{-1}\begin{pmatrix} Q_0 \\ (X_1'M_0X_1)^{-1}X_1'M_0 \end{pmatrix},$$

where

$$Q_0 = (X_0'X_0)^{-1}X_0'\left(I_n - X_1(X_1'M_0X_1)^{-1}X_1'M_0\right),$$

and hence

$$\begin{aligned} Q_* &= (Z_*'Z_*)^{-1}Z_*'Z_2 = \Delta_*^{-1}\begin{pmatrix} Q_0X_2\Delta_2\Psi^{-1/2} \\ \breve{\Delta}_1\breve{Q} \end{pmatrix}, \\ \hat{\gamma}_{*r} &= (Z_*'Z_*)^{-1}Z_*'y = \Delta_*^{-1}\begin{pmatrix} Q_0y \\ \breve{\Delta}_1\hat{\gamma}_{1r} \end{pmatrix}, \end{aligned}$$

leading to the WALS estimator of β_*:

$$\tilde{\beta}_* = \Delta_*(\hat{\gamma}_{*r} - Q_*\tilde{\gamma}_2) = \begin{pmatrix} (X_0'X_0)^{-1}X_0'(y - X_1\tilde{\beta}_1 - X_2\tilde{\beta}_2) \\ \tilde{\beta}_1 \end{pmatrix} = \begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix}.$$

This concludes the proof.

Proof of Proposition 4: Let $\xi_1 > 0$ and $\xi_2 > 0$. Then,

$$\left(\xi_1^{-1}J_T + \xi_2^{-1}(I_T - J_T)\right)\left(\xi_1 J_T + \xi_2(I_T - J_T)\right) = I_T$$

and

$$\left(\xi_1^{-1/2}J_T + \xi_2^{-1/2}(I_T - J_T)\right)\left(\xi_1^{-1/2}J_T + \xi_2^{-1/2}(I_T - J_T)\right)$$
$$= \xi_1^{-1}J_T + \xi_2^{-1}(I_T - J_T).$$

Setting $\xi_1 = T\alpha_1 + 1$ and $\xi_2 = 1$ completes the proof.

References

Baltagi, B. H. (2021). *Econometric Analysis of Panel Data*, 6th edition. Springer, New York.

Baltagi, B. H., and Q. Li (1991). A transformation that will circumvent the problem of autocorrelation in an error-component model. *Journal of Econometrics*, 48, 385–393.

Baltagi, B. H., and P. X. Wu (1999). Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory*, 15, 814–823.

Belloni, A., V. Chernozhukov, and C. Hansen (2014a). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81, 608–650.

Belloni, A., V. Chernozhukov, and C. Hansen (2014b). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28, 29–50.

Bhadra, A., J. Datta, N. G. Polson, and B. Willard (2017). Horseshoe regularization for feature subset selection.
ArXiv preprint arXiv:1702.07400.

Bhargava, A., L. Franzini, and W. Narendranathan (1982). Serial correlation and the fixed effects model. *Review of Economic Studies*, 49, 533–549.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97, 465–480.

Claeskens, G., and N. L. Hjort (2008). *Model Selection and Model Averaging*. Cambridge University Press, New York.

Danilov, D. (2005). Estimation of the mean of a univariate normal distribution when the variance is not known. *Econometrics Journal*, 8, 277–291.

De Luca, G., and J. R. Magnus (2025). *Weighted-Average Least Squares Model Averaging*, monograph, in progress.

De Luca, G., J. R. Magnus, and F. Peracchi (2022). Sampling properties of the Bayesian posterior mean with an application to WALS estimation. *Journal of Econometrics*, 230, 299–317.

De Luca, G., J. R. Magnus, and F. Peracchi (2023). Weighted-average least squares (WALS): Confidence and prediction intervals. *Computational Economics*, 61, 1637–1664.

De Luca, G., J. R. Magnus, and F. Peracchi (2024). Bayesian estimation of the normal location model: A non-standard approach. EIEF Working Paper 24/03. https://www.eief.it.

Donohue, J. J., and S. D. Levitt (2001). The impact of legalized abortion on crime. *Quarterly Journal of Economics*, 116, 379–420.

Donohue, J. J., and S. D. Levitt (2020). The impact of legalized abortion on crime over the last two decades. *American Law and Economics Review*, 22, 241–302.

Hahn, P. R., C. M. Carvalho, D. Puelz, and J. He (2018). Regularization and confounding in linear regression for treatment effect estimation. *Bayesian Analysis*, 13, 163–182.

Leeb, H., and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21, 21–59.

Leeb, H., and B. M. Pötscher (2006). Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econometric Theory*, 22, 69–97.

Magnus, J. R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal*, 5, 225–236.

Magnus, J. R., and G. De Luca (2016). Weighted-average least squares: A review. *Journal of Economic Surveys*, 30, 117–148.

Magnus, J. R., and J. Durbin (1999). Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica*, 67, 639–643.

Magnus, J. R., O. Powell, and P. Prüfer (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154, 139–153.

Magnus, J. R., A. T. K. Wan, and X. Zhang (2011). Weighted average least squares estimation with nonspherical disturbances and an application to the Hong Kong housing market. *Computational Statistics & Data Analysis*, 55, 1331–1341.

McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, 200, 17–25.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29, 46–75.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46, 69–85.

Saleh, A. K. M. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation With Applications*. Wiley, New York.

Steel, M. F. J. (2020). Model averaging and its use in economics. *Journal of Economic Literature*, 58, 644–719.

Acknowledgements To follow.

Competing Interests The authors have no conflicts of interest to declare that are relevant to the content of this chapter.