university of groningen

faculty of economics and business

2020002-EEF

Refining Clustered Standard Errors with Few Clusters

January 2020

Gianmaria Niccodemi
Rob Alessie
Viola Angelini
Jochen Mierau
Tom Wansbeek

SOM is the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:
- Economics, Econometrics and Finance
- Global Economics & Management
- Innovation & Organization
- Marketing
- Operations Management & Operations Research
- Organizational Behaviour

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 9090/7068/3815

www.rug.nl/feb/research

Refining Clustered Standard Errors with Few Clusters

Gianmaria Niccodemi
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance
g.niccodemi@rug.nl

Rob Alessie
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance

Viola Angelini
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance

Jochen Mierau
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance

Tom Wansbeek
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance

Refining clustered standard errors with few clusters[*†]

Gianmaria Niccodemi[‡] Rob Alessie Viola Angelini Jochen Mierau

Tom Wansbeek

Faculty of Economics and Business, University of Groningen

Abstract

We introduce efficient formulas that dramatically decrease the computational time of CR2VE and CR3VE, the cluster-robust estimators of standard errors with few clusters, and of the Imbens and Kolesar (2016) degrees of freedom. We also introduce CR3VE-λ, an estimator that is unbiased under more general conditions than CR3VE as it takes cluster unbalancedness into account. We illustrate these refinements by empirical simulations.

1 Introduction

In linear regressions with clustered data it is common practice to estimate the variance of the estimated parameters with CRVE, the cluster-robust estimator introduced by Liang and Zeger (1986) as a generalization of the White's (1980) heteroscedastic-robust estimator. Unbiasedness of CRVE relies on the assumption that the number of clusters tends to infinity. With few clusters and error term correlated within cluster CRVE leads to downward biased standard errors and thus

[*]We are grateful to Nick Koning, Erik Meijer, Douglas Miller and Ulrich Schneider for helpful comments and suggestions. We are also grateful to conference audiences at NESG 2019 in Amsterdam, KVS New Papers Session in The Hague and RSS 2019 in Belfast, and to the internal seminar audience at the University of Groningen.

[†]The CPS 2012 data used in this paper is obtainable from https://cps.ipums.org/cps/. All the authors declare that they have no financial interests that relate to the research described in this paper.

[‡]Corresponding author. e-mail: g.niccodemi@rug.nl; telephone: +31 50 36 37018; address: Nettelbosje 2, 9747 AE, Groningen, Netherlands.

misleading inference on the estimated parameters. Moulton (1986, 1990) and Cameron and Miller (2015) point out that this issue is particularly relevant for regressors that are constant within cluster such as policy variables that are only implemented in certain regions or states. An additional issue for inference on a single estimated parameter is that, under the null hypothesis and with few clusters, the distribution of the test statistic is unknown and not approximable to the standard normal.

Bell and McCaffrey (2002) propose to improve the inference on the single parameter by (i) reducing the bias of CRVE with either CR2VE, also known as BRL (bias reduced linearization), or CR3VE, both based on transformed OLS residuals, and by (ii) approximating the distribution of the test statistic with the t-distribution with degrees of freedom (DOF) that are data-determined and regressor-specific. Imbens and Kolesar (2016) develop a more refined version of the data-determined regressor-specific DOF used by Bell and McCaffrey (2002), IK from here on.

Unfortunately, these methods have drawbacks that are particularly relevant for empirical research. First, CR2VE, CR3VE and the IK may be computationally demanding as they are based on the computation of the inverse (CR3VE) and the inverse square root (CR2VE and the IK) of square matrices of order equal to the number of observations per cluster. Second, if the few clusters are highly unbalanced CR3VE standard errors may be too conservative and may lead to underrejection of a true null hypothesis.

In view of these issues, this paper presents some results that are particularly meant for empirical researchers who wish to estimate a linear model on cross-sectional data clustered in few clusters. We show how to compute CR2VE, CR3VE and the IK efficiently, regardless of the size of the clusters, by inverting matrices of order equal to the number of regressors only. Moreover, we introduce CR3VE-λ, a cluster-robust variance estimator that is identical to CR3VE in case of balanced clusters but, in case of unbalanced clusters, takes the difference in cluster sizes into account to make the computed standard errors closer to unbiasedness. Through simulations we show that, with high unbalancedness of the few clusters and using the t(IK) distribution, CR3VE-λ leads to better inference than CR3VE. Moreover, we show that our efficient formulas produce high gains in terms of computational time: for example, more than three hours can be saved for the computation of CR2VE and CR3VE on a standard machine using a dataset with 10 clusters and 5,000 observations per cluster.

The remaining of this paper is organized as follows. In Section 2 we discuss basic theory on CRVE, CR2VE and CR3VE. In Section 3 we introduce CR3VE-λ. In Sections 4 and 5 we introduce the formulas to compute CR2VE, CR3VE (and CR3VE-λ) and the IK efficiently. In Section 6 we illustrate and test the performance of CRVE, CR2VE, CR3VE and CR3VE-λ to compute standard errors with few clusters by Monte Carlo simulations. In Section 6 we also show the computational time gain from our efficient formulas for CR2VE and CR3VE using data with different number and size of clusters. In Section 7 we conclude the paper with recommendations for empirical researchers.

For all the computations and the empirical illustrations we use Stata/SE 15.0, as Stata is the statistical software most used by empirical researchers. The Stata do-file that can be used with any cross-sectional dataset for computing standard errors based on the discussed methods and the Stata do-files to replicate the experiments and the simulated datasets are available upon request.

2 Basic theory: CRVE, CR2VE and CR3VE

Define the regression model with k regressors $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and consider observations that can be grouped into $i = 1, \ldots, c$ clusters of size n_i, $\sum_i n_i = n$, and write, for the i-th cluster

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

with $\mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\varepsilon}_i) = \mathbf{V}_i$. The $\mathbf{V}_i$'s are collected in the block-diagonal matrix $\mathbf{V}$. After OLS we have

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{V}_i\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{1}$$

The "classical", non-robust estimator of (1) is biased and it will usually underestimate the true variance since

$$\mathrm{E}[\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})] = \frac{\mathrm{tr}[\mathbf{M}\mathbf{V}]}{n-k}(\mathbf{X}'\mathbf{X})^{-1}, \tag{2}$$

where $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. To avoid the bias, an obvious estimator is the cluster-robust variance estimator (CRVE) based on OLS residuals per cluster $\hat{\boldsymbol{\varepsilon}}_i$

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \mathbf{X}_i'\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{3}$$

This estimator, which is introduced by Liang and Zeger (1986) and generalizes White (1980), is consistent when the number of clusters goes to infinity. In case of few clusters asymptotics will be a poor guide. Therefore we consider its bias instead.

Let $\mathbf{S}_i$ be the $n \times n_i$ matrix that selects the columns of $\mathbf{M}$ corresponding to cluster i and define

$$\begin{aligned} \mathbf{L}_i &\equiv \mathbf{M}\mathbf{S}_i \\ \mathbf{H}_i &\equiv \mathbf{S}_i'\mathbf{M}\mathbf{S}_i = \mathbf{I}_i - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i', \end{aligned}$$

where $\mathbf{I}_i$ is the $n_i \times n_i$ identity matrix.[1] There holds $\mathbf{H}_i = \mathbf{L}_i'\mathbf{L}_i$ since $\mathbf{M}$ is idempotent and symmetric. With $\hat{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\varepsilon}}_i = \mathbf{L}_i'\boldsymbol{\varepsilon}$, we have

$$\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i') = \mathbf{L}_i'\mathbf{V}\mathbf{L}_i \neq \mathbf{V}_i,$$

so

$$\mathrm{E}[\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})] = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{L}_i'\mathbf{V}\mathbf{L}_i\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1} \neq \mathrm{var}(\hat{\boldsymbol{\beta}}).$$

To reduce the bias, we consider a variance estimator based on transformed residuals

$$\tilde{\boldsymbol{\varepsilon}}_i \equiv \mathbf{A}_i\hat{\boldsymbol{\varepsilon}}_i,$$

for some $\mathbf{A}_i$ to be chosen. Then

$$\mathrm{E}[\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})] = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{A}_i\mathbf{L}_i'\mathbf{V}\mathbf{L}_i\mathbf{A}_i'\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1}.$$

From (1), unbiasedness requires the $\mathbf{A}_i$ to be such that $\mathbf{A}_i\mathbf{L}_i'\mathbf{V}\mathbf{L}_i\mathbf{A}_i' = \mathbf{V}_i$ for all i uniformly in the $\mathbf{V}_i$. This is infeasible and therefore we consider two second-best solutions.

The first second-best solution is to consider the case where there are no cluster effects, $\mathbf{V}_i = \sigma^2\mathbf{I}_i$ for all i, and make the estimator unbiased for this case. Then $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i') = \mathbf{L}_i'\mathbf{V}\mathbf{L}_i = \sigma^2\mathbf{L}_i'\mathbf{L}_i = \sigma^2\mathbf{H}_i$ and consequently

$$\mathrm{E}[\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{A}_i\mathbf{H}_i\mathbf{A}_i'\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{4}$$

The variance estimator is unbiased if $\mathbf{A}_i\mathbf{H}_i\mathbf{A}_i' = \mathbf{I}_i$ and consequently we choose $\mathbf{A}_i = \mathbf{H}_i^{-1/2}$. This estimator, introduced by Bell and McCaffrey (2002) and extensively discussed by Cameron and Miller (2015), is known as both CR2VE and BRL.

[1] For the sake of readability we write $\mathbf{I}_i$ instead of $\mathbf{I}_{n_i}$. Likewise we will indicate an n_i-vector of ones as $\boldsymbol{\iota}_i$ and an $n_i \times n_i$-matrix of ones as $\mathbf{J}_i$.

The other second-best solution is based on the idea that the elements in $\mathbf{M}$ outside the blocks on the diagonal may be small and therefore negligible. Then $\mathbf{L}_i$ can be approximated by a matrix with $\mathbf{H}_i$ as its ith block and zeros outside this block. Then $\mathbf{L}_i'\mathbf{V}\mathbf{L}_i = \mathbf{H}_i\mathbf{V}_i\mathbf{H}_i$ and choosing $\mathbf{A}_i = \mathbf{H}_i^{-1}$ leads, when scaled by a factor $(c-1)/c$, to an estimator that is approximately unbiased when there are no cluster effects. This estimator is introduced by Bell and McCaffrey (2002) and discussed by Cameron and Miller (2015) and it is known as CR3VE.

To analyze the bias of CR3VE we scale (4) by $(c-1)/c$ and use

$$\mathbf{A}_i\mathbf{H}_i\mathbf{A}_i = \mathbf{H}_i^{-1} = \mathbf{I}_i + \mathbf{X}_i(\mathbf{X}'\mathbf{X} - \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'$$

to obtain

$$\mathrm{E}[\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})] = \frac{c-1}{c}\sigma^2\left((\mathbf{X}'\mathbf{X})^{-1} + \sum_i (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X} - \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\right). \tag{5}$$

When clusters are balanced and have the same covariance structure there holds $\mathbf{X}_i'\mathbf{X}_i = \frac{1}{c}\mathbf{X}'\mathbf{X}$ for all i, and (5) reduces to $\mathrm{E}[\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}})] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Therefore, in case of balanced clusters, CR3VE with the correction factor $(c-1)/c$ is unbiased.

3 From CR3VE to CR3VE-λ

We propose a different scaling factor than $(c-1)/c$ for CR3VE in the more general case of unbalanced clusters that still have the same covariance structure. Define $\pi_i \equiv n_i/n$ for cluster i. Then we have $\mathbf{X}_i'\mathbf{X}_i = \pi_i\mathbf{X}'\mathbf{X}$ and in (5)

$$(\mathbf{X}'\mathbf{X})^{-1} + \sum_i (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X} - \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1} = \lambda(\mathbf{X}'\mathbf{X})^{-1},$$

with

$$\lambda \equiv 1 + \sum_i \frac{\pi_i^2}{1-\pi_i}.$$

There holds $\lambda \geq c/(c-1)$, with equality in case of balanced clusters. To see this, let $\boldsymbol{\pi} \equiv (\pi_1, \ldots, \pi_c)'$ and $\boldsymbol{\Pi} \equiv \mathrm{diag}(\boldsymbol{\pi})$, and let

$$\begin{aligned} \mathbf{a} &\equiv (\mathbf{I}_c - \boldsymbol{\Pi})^{-1/2}\boldsymbol{\pi} \\ \mathbf{b} &\equiv (\mathbf{I}_c - \boldsymbol{\Pi})^{1/2}\boldsymbol{\iota}_c \end{aligned}$$

so $\mathbf{a}'\mathbf{a} = \boldsymbol{\pi}'(\mathbf{I}_c - \boldsymbol{\Pi})^{-1}\boldsymbol{\pi}$, $\mathbf{b}'\mathbf{b} = \boldsymbol{\iota}_c'(\mathbf{I}_c - \boldsymbol{\Pi})\boldsymbol{\iota}_c$, and $\mathbf{a}'\mathbf{b} = 1$. Since $(\mathbf{a}'\mathbf{b})^2 \leq \mathbf{a}'\mathbf{a}\ \mathbf{b}'\mathbf{b}$ there holds

$$\sum_i \frac{\pi_i^2}{1-\pi_i} = \boldsymbol{\pi}'(\mathbf{I}_c - \boldsymbol{\Pi})^{-1}\boldsymbol{\pi} \geq \frac{1}{\boldsymbol{\iota}'(\mathbf{I}_c - \boldsymbol{\Pi})\boldsymbol{\iota}} = \frac{1}{c-1},$$

so $\lambda - 1 \geq 1/(c-1)$ or $\lambda \geq c/(c-1)$. This suggests that $1/\lambda$ may be a better scaling factor than $(c-1)/c$. We denote this estimator, which is unbiased under more general conditions than CR3VE, by CR3VE-λ.

4 Efficient computation of CR2VE, CR3VE, CR3VE-λ with $\mathbf{H}_i$

CR2VE and CR3VE are based on $\mathbf{H}_i^a$, with $a = -1/2$ and $a = -1$, respectively. Especially with large n_i it is desirable to exploit the structure of $\mathbf{H}_i$ for the computations. We do so through the following result, that allows for reducing the computing and storage requirements to be just $O(n_i)$ instead of $O(n_i^2)$ for storage and $O(n_i^3)$ for inversion.[2] Let $\mathbf{R}$ be a matrix of "large" number of rows ℓ and "small" number of columns s, $\ell \geq s$, and let $\mathbf{R}$ have full column rank and satisfy $\mathbf{R}'\mathbf{R} \leq \mathbf{I}_s$. Then

$$\mathbf{R}'(\mathbf{I}_\ell - \mathbf{R}\mathbf{R}')^a = (\mathbf{I}_s - \mathbf{R}'\mathbf{R})^a\mathbf{R}' \tag{6}$$

for any a. To see this, take the singular value decomposition $\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{T}'$, with $\boldsymbol{\Lambda}$ diagonal, $\mathbf{T}$ square orthonormal, and $\mathbf{U}$ having orthonormal columns. Then both sides of (6) appear to be equal to $\mathbf{T}\boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Lambda}^2)^a\mathbf{U}'$.

Now, let $\hat{\mathbf{s}}_i \equiv \mathbf{X}_i'\hat{\boldsymbol{\varepsilon}}_i$. Then the right-hand side of (3) can be written as $(\mathbf{X}'\mathbf{X})^{-1}[\sum_i \hat{\mathbf{s}}_i\hat{\mathbf{s}}_i'](\mathbf{X}'\mathbf{X})^{-1}$. With CR2VE ($a = -1/2$), CR3VE ($a = -1$) and CR3VE-$\lambda$ ($a = -1$), $\hat{\mathbf{s}}_i$ has to be replaced by $\tilde{\mathbf{s}}_i \equiv \mathbf{X}_i'\mathbf{H}_i^a\hat{\mathbf{s}}_i$, still with scaling to be added for CR3VE and CR3VE-λ. Define

$$\mathbf{R}_i \equiv \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1/2}, \tag{7}$$

so $\mathbf{X}_i' = (\mathbf{X}'\mathbf{X})^{1/2}\mathbf{R}_i'$ and $\mathbf{H}_i = \mathbf{I}_i - \mathbf{R}_i\mathbf{R}_i'$. Then from (6)

$$\begin{aligned}\tilde{\mathbf{s}}_i &= (\mathbf{X}'\mathbf{X})^{1/2}\mathbf{R}_i'(\mathbf{I}_i - \mathbf{R}_i\mathbf{R}_i')^a\hat{\boldsymbol{\varepsilon}}_i \\ &= (\mathbf{X}'\mathbf{X})^{1/2}(\mathbf{I}_k - \mathbf{R}_i'\mathbf{R}_i)^a(\mathbf{X}'\mathbf{X})^{-1/2}\hat{\mathbf{s}}_i.\end{aligned}$$

[2] Le Gall (2014) gives the best-known lower bound of $O(n^{2.373})$. This is mainly of theoretical value and it holds for the optimized Coppersmith-Winograd algorithm.

So the computations to obtain $\tilde{\mathbf{s}}_I$ involve only matrices of order $k \times k$, which is $O(1)$ in n_i given $\mathbf{R}_i'\mathbf{R}_i$, $\mathbf{X}'\mathbf{X}$ and $\hat{\mathbf{s}}_i$; all three are computable in $O(n_i)$. This essentially simplifies the computation of CR2VE, CR3VE and CR3VE-λ. In Appendix A we summarize the formulas for CRVE, CR2VE, CR3VE and CR3VE-λ.

5 Efficient computation of the Imbens and Kolesar degrees of freedom

Define $\hat{\beta}_r$ the estimated coefficient of the rth regressor, $r = 1, \ldots, k$. With few clusters the distribution under the null of the test statistic for inference on $\hat{\beta}_r$ is unknown and not approximable to $N(0, 1)$. It is common practice in empirical research to use the t-distribution with $(c - 1)$ DOF or, more recently, with the IK developed by Imbens and Kolesar (2016) and based on $\mathbf{H}_i^{-1/2}$.

Define the $n \times c$ matrix $\mathbf{F}_r$ with ith column equal to

$$\mathbf{F}_{ri} = \mathbf{G}_i \mathbf{e}_r, \tag{8}$$

where $\mathbf{G}_i = \mathbf{L}_i \mathbf{H}_i^{-1/2} \mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1}$ and $\mathbf{e}_r$ is a k-vector with rth element equal to 1 and any other elements equal to 0. Consider the random effect parametrization of $\mathbf{V} = \sigma^2 \mathbf{I}_n + \theta^2 \mathbf{D}\mathbf{D}'$, where $\mathbf{D} \equiv \text{diag}(\boldsymbol{\iota}_i)$. Then the IK for regressor r are

$$\text{IK}_r = \frac{(\sum_i \kappa_i)^2}{\sum_i \kappa_i^2}, \tag{9}$$

where κ_i are the eigenvalues of $\mathbf{F}_r'\hat{\mathbf{V}}\mathbf{F}_r \equiv \hat{\sigma}^2 \mathbf{F}_r'\mathbf{F}_r + \hat{\theta}^2 \mathbf{F}_r'\mathbf{D}\mathbf{D}'\mathbf{F}_r$ and $\hat{\sigma}^2$ and $\hat{\theta}^2$ can be obtained from a random effect estimation.

Based on (6) and with $\mathbf{R}$ as defined in (7) we derive the efficient formula for $\mathbf{G}_i$ as

$$\begin{aligned}
\mathbf{G}_i &= \mathbf{L}_i \mathbf{H}_i^{-1/2} \mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1} \\
&= \mathbf{L}_i (\mathbf{I}_i - \mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i')^{-1/2} \mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1} \\
&= \mathbf{L}_i (\mathbf{I}_i - \mathbf{R}_i \mathbf{R}_i')^{-1/2} \mathbf{R}_i (\mathbf{X}'\mathbf{X})^{-1/2} \\
&= \mathbf{L}_i \mathbf{R}_i (\mathbf{I}_k - \mathbf{R}_i'\mathbf{R}_i)^{-1/2} (\mathbf{X}'\mathbf{X})^{-1/2} \\
&= [\mathbf{S}_i - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'] \mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1/2} [\mathbf{I}_k - (\mathbf{X}'\mathbf{X})^{-1/2} \mathbf{X}_i'\mathbf{X}_i (\mathbf{X}'\mathbf{X})^{-1/2}]^{-1/2} (\mathbf{X}'\mathbf{X})^{-1/2} \\
&\equiv \mathbf{S}_i \mathbf{X}_i \mathbf{W}_i - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i'\mathbf{X}_i \mathbf{W}_i,
\end{aligned}$$

where $\mathbf{S}_i\mathbf{X}_i\mathbf{W}_i$ is the $n \times k$ matrix with block that corresponds to cluster i equal to $\mathbf{X}_i\mathbf{W}_i$ and all the other rows equal to $\mathbf{0}$, and where $\mathbf{W}_i = (\mathbf{X}'\mathbf{X})^{-1/2}[\mathbf{I}_k - (\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}_i'\mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1/2}]^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}$.

6 Empirical illustration

Table 1: Rejection rates *policy* from 20,000 MC replications

		No. of states				
Method	Distribution	6	10	14	20	50
Unclustered s.e.	$t(c-1)$	30.1	36.5	40.0	41.2	44.1
CRVE	$t(c-1)$	14.0	10.6	9.7	8.3	6.8
CR2VE	$t(c-1)$	7.9	7.0	6.9	6.6	6.1
CR3VE	$t(c-1)$	5.3	5.3	5.2	5.4	5.6
CR3VE-λ	$t(c-1)$	5.8	5.7	5.6	5.6	5.7

Rejection rates, in percentage, of the true null hypothesis on the fake *policy* variable from 20,000 MC replications for different methods to compute standard errors. Ideal rejection rates are equal to 5%. 20% observations within sampled states are randomly sampled with replacement. The 6, 10, 14, 20, 30, 50 states are randomly sampled with replacement. $t(c-1)$ distribution is used for inference. Stata/SE 15.0 is used for simulations.

Cameron and Miller (2015) point out that inference on constant within-cluster variables is problematic with few clusters, even with a low intra-cluster correlation of the error term. Both a low number of clusters and a low intra-cluster correlation can be typically found in cross-sectional data of individuals clustered at some geographical levels. Using such cross-sectional data we run Monte Carlo (MC) simulations to test inference based on unclustered standard errors, CRVE, and CR2VE, CR3VE and CR3VE-λ computed efficiently (see Section 4). According to Cameron and Miller (2015), at least the $t(c-1)$ distribution or the more effective t(IK) distribution should be used for inference on the single estimated parameter. We use both for our simulations and we use the efficient formula for the computation of the IK (see Section 5). Section 6.1 concludes the empirical illustration with a discussion on the computational time gain from our efficient formulas

Table 2: Rejection rates *policy* from 20,000 MC replications

		No. of states			
Method	Distribution	6	6 hu_1	6 hu_2	6 hu_3
Unclustered s.e.	t(IK)	21.8	23.0	47.4	78.5
CRVE	t(IK)	9.6	8.8	15.8	50.8
CR2VE	t(IK)	5.3	3.4	4.4	12.3
CR3VE	t(IK)	3.3	2.0	1.5	0.9
CR3VE-λ	t(IK)	3.8	2.8	3.2	4.6
	mean(IK)	3.3	2.2	2.5	3.1
	$1/\lambda$		0.69	0.61	0.41

Rejection rates, in percentage, of the true null hypothesis on the fake *policy* variable from 20,000 MC replications for different methods to compute standard errors. Ideal rejection rates are equal to 5%. 20% observations within sampled states are randomly sampled with replacement. The 6 states are randomly sampled with replacement. For all replications, the 6 highly unbalanced hu_1, hu_2 and hu_3 states are the 3 with most observations and the 3 with least observations, the 2 with most observations and the 4 with least observations, and the 1 with most observations and the 5 with least observations, respectively. t(IK) distribution is used for inference. The variance components for computing IK are estimated with restricted maximum likelihood (`mixed,reml` or `xtmixed,reml` command in Stata). Stata/SE 15.0 is used for simulations.

for CR2VE and CR3VE with respect to the ones introduced by Bell and McCaffrey (2002).

In our empirical illustration we perform the same MC set-up as in Cameron and Miller (2015). We use the same dataset CPS 2012 which consists of 51 clusters, namely the 50 American States and the District of Columbia, and we define the same model for individual h in the sampled cluster $i = 1, \ldots, c$

$$\ln(wage)_{hi} = \beta_0 + \beta_1 educ_{hi} + \beta_2 age_{hi} + \beta_3 age_{hi}^2 + \beta_4 policy_i + \varepsilon_{hi}, \tag{10}$$

where *policy* is a fake policy variable randomly assigned to $c/2$ sampled clusters and constant within each cluster. The clusters are unbalanced and the number of observations per cluster is reported in Table B.1 in Appendix B.

We run 5 sets of 20,000 MC replications using a random sample with replacement of $c = 6, 10, 14, 20, 50$ clusters. In order to preserve the unbalancedness of the clusters, we randomly sample with replacement 20% of the observations within each sampled cluster. In each simulation we test the true null hypothesis $H_0 : \beta_4 = 0$ at the 5% level and thus we expect the standard errors of *policy* to lead to rejection of the true null hypothesis H_0 in 5% of the replications. Rejection rates using for inference the $t(c-1)$ distribution are reported in Table 1. Inference based on unclustered standard errors or, with few clusters, CRVE is clearly misleading. The rejection rates of CR2VE and CR3VE computed with our formulas are, as expected, in line with those reported by Cameron and Miller (2015).[3] CR3VE-λ rejection rates do not differ much from those of CR3VE but this might depend on the clusters being not highly unbalanced. We report the rejection rates for the experiment with 6 clusters using for inference the more effective t(IK) distribution in column 3 of Table 2. As expected, the rejection rates of all methods decrease using a distribution with, on average, 3.3 DOF instead of 5, with CR3VE-λ rejection rate closer to 5% than CR3VE rejection rate.

To test CR3VE-λ with higher unbalancedness of clusters we run three more empirical illustrations of 20,000 MC replications on model (10). In the first (hu_1) we use only the 3 states with most individuals and the 3 states with least individuals, in the second (hu_2) we use only the 2 states with most individuals and the 4 states with least individuals and in the third (hu_3) we use only the state with most individuals and the 5 states with least individuals (see Table B.1 in Appendix B for the number of observed individuals in the CPS 2012 dataset). Similarly to the first empirical illustration, we sample with replacement 20% of the observations within each of these states. Rejection rates of hu_1, hu_2 and hu_3 using for inference the t(IK) distribution are reported in Table 2. While the scaling factor for CR3VE is constant and equal to 0.83, the scaling factor

[3] An obvious advantage of using our formulas for CR2VE, CR3VE and CR3VE-λ is that we are able to run 20,000 replications for each number of clusters in short time. Cameron and Miller (2015), for the same experiments, run only 4,000 replications for 6 and 10 clusters and 1,000 replications for 20 clusters or more, presumably due to the time-consuming inefficient formulas.

for CR3VE-λ decreases from hu_1 to hu_2 and from hu_2 to hu_3, making the standard errors based on CR3VE-λ closer to unbiasedness than the standard errors based on CR3VE.

As expected the improvement based on CR3VE-λ is particularly relevant with high unbalancedness. An indicator of high unbalancedness might be the effective number of clusters developed by Carter, Schnepel, and Steigerwald (2017). If the decrease in the effective number of clusters with respect to the nominal one depends on higher unbalancedness in cluster size then CR3VE-λ should lead to less conservative and thus less upward biased standard errors than CR3VE.

6.1 Computational time gain for CR2VE and CR3VE

Table 3: Time in seconds for CR2VE and CR3VE using efficient and inefficient formulas

		No. of observations per cluster				
CR2VE+CR3VE	No. of clusters	1000	2000	3000	4000	5000
Efficient	6	1	1	1	1	1
Inefficient	6	44	368	1371	3569	6943
Efficient	10	1	1	1	1	1
Inefficient	10	67	599	2214	5630	11 568

Total computational time of CR2VE and CR3VE using the formulas reported in Section 4 (efficient) and in Cameron and Miller (2015) (inefficient). The computations are run using Stata/SE 15.0 on the following machine: Intel(R) Core(TM) i3-4130 CPU @ 3.40GHz, RAM: 8,00 GB, Windows 7.

We report in Table 3 the computational time of our efficient formulas for CR2VE and CR3VE and of the equivalent, but inefficient, CR2VE and CR3VE as introduced by Bell and McCaffrey (2002). This computational time refers to CR2VE and CR3VE estimated together on a standard machine. We run these computations on simulated data with 51 balanced clusters and 5000 observations per cluster. The data generating process is $\ln(wage)_{hi} = 0.7495 + 0.0844age_{hi} - 0.0009age_{hi}^2 + u_i + e_{hi}$, where $age \sim U\{18, 65\}$, $u_i \sim N(0, \theta^2)$ is constant within cluster i and $e_{hi} \sim N(0, \sigma^2)$, and where age_{hi}, u_i and e_{hi} are mutually independent. We set $\theta^2 = 9.5818 \times 10^{-3}$ and $\sigma^2 = 0.3489$. The parameters of the data generating process and of the u_i and e_{hi} distributions are chosen from a random effects regression on the CPS 2012 dataset, using all the 51 clusters and

all the observations. Based on this data we define the model

$$\ln(wage)_{hi} = \beta_0 + \beta_1 age_{hi} + \beta_2 age_{hi}^2 + \beta_3 policy_i + \varepsilon_{hi}, \tag{11}$$

where *policy* is a fake policy variable randomly assigned to half of the clusters.

We sample 6 and 10 clusters, and 1000, 2000, 3000, 4000 and 5000 observations within each cluster from this simulated data. We compute the clustered standard errors based on CR2VE and CR3VE with these different samples. The computations of the inefficient CR2VE and CR3VE take up to more than three hours for 10 clusters. This depends on the fact that the inefficient formulas invert matrices of order $n_i \times n_i$ and thus the computational time increases with cluster size n_i. Oppositely, as shown in Section 4 the efficient formulas invert matrices of order $k \times k$ that does not depend on the cluster size n_i, where $k = 4$ is the number of regressors in model (11).

7 Conclusion

We have illustrated results that might be particularly useful for empirical researchers who wish to compute clustered standard errors in case of few clusters. First, CR3VE-λ is unbiased under more general conditions than CR3VE as it takes cluster unbalancedness into account. Second, the efficient formulas for CR2VE, CR3VE (and CR3VE-λ) and the IK invert much lower-order matrices than the standard formulas. Remarkably, this order does not depend on the size of the clusters. We recommend the empirical researcher to use the efficient formulas for CR2VE and CR3VE (and CR3VE-λ) in case of large cluster sizes as this saves a remarkable amount of time for computation. Moreover, based on the empirical results, we recommend to use CR3VE-λ rather than CR3VE especially in case of few highly unbalanced clusters.

The Stata do-file that can be used with any cross-sectional dataset for computing standard errors based on the discussed methods and the Stata do-files to replicate the experiments and the simulated datasets are available upon request.

References

Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*, 169-179.

Cameron, A., & Miller, D. (2015). A practitioner‘s guide to cluster-robust inference. *Journal of Human Resources*, *50*, 317-372.

Carter, A. V., Schnepel, K. T., & Steigerwald, D. G. (2017). Asymptotic behavior of a t-test robust to cluster heterogeneity. *Review of Economics and Statistics*, *99*(4), 698–709.

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, *98*(4), 701–712.

Le Gall, F. (2014). Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation* (pp. 296–303).

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics*, *32*(3), 385–397.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. *Review of Economics and Statistics*, *72*(2), 334–338.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

Appendix A CRVE, CR2VE, CR3VE and CR3VE-λ in a nutshell

Define the matrix of observations $\mathbf{X}$ of order $n \times k$ and the linear model for cluster $i = 1, \ldots, c$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{X}_i$ is a matrix of order $n_i \times k$, and where $\mathrm{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\varepsilon}_i) = \mathbf{V}_i$. Define the OLS residuals $\hat{\boldsymbol{\varepsilon}}_i$. The general expression for the cluster-robust estimator of $\mathrm{var}(\hat{\boldsymbol{\beta}})$ is

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \mathbf{X}_i'\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i'\mathbf{X}_i\right)(\mathbf{X}'\mathbf{X})^{-1},$$

where $\tilde{\boldsymbol{\varepsilon}}_i$ are a transformation of OLS residuals to be specified. CRVE simply uses $\tilde{\boldsymbol{\varepsilon}}_i = \hat{\boldsymbol{\varepsilon}}$. CR2VE uses $\tilde{\boldsymbol{\varepsilon}}_i = (\mathbf{I}_i - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i')^{-1/2}\hat{\boldsymbol{\varepsilon}}_i$, while CR3VE and CR3VE-$\lambda$ use $\tilde{\boldsymbol{\varepsilon}}_i = g[(\mathbf{I}_i - \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i')^{-1}\hat{\boldsymbol{\varepsilon}}_i]$, where $g = [(c-1)/c]^{1/2}$ for CR3VE and $g = \{1 + [\sum_i (n_i/n)^2/(1 - n_i/n)]\}^{-1/2}$ for CR3VE-λ. In case of balanced clusters CR3VE and CR3VE-λ are identical. Only CRVE requires $c \to \infty$ which, in empirical applications, means that the number of clusters has to be sufficiently large.

CR2VE, CR3VE and CR3VE-λ can be computed efficiently with the inversion of matrices of order $k \times k$ instead of $n_i \times n_i$. Define $\hat{\mathbf{s}}_i = \mathbf{X}_i'\hat{\boldsymbol{\varepsilon}}_i$, $\mathbf{R}_i' = \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1/2}$ and the cluster robust variance estimator

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \tilde{\mathbf{s}}_i\tilde{\mathbf{s}}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}.$$

Then to compute CR2VE we use $\tilde{\mathbf{s}}_i = [(\mathbf{X}'\mathbf{X})^{1/2}(\mathbf{I}_k - \mathbf{R}_i'\mathbf{R}_i)^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}]\hat{\mathbf{s}}_i$, and to compute CR3VE and CR3VE-λ we use $\tilde{\mathbf{s}}_i = g[(\mathbf{X}'\mathbf{X})^{1/2}(\mathbf{I}_k - \mathbf{R}_i'\mathbf{R}_i)^{-1}(\mathbf{X}'\mathbf{X})^{-1/2}\hat{\mathbf{s}}_i]$.

Appendix B Additional tables

Table B.1: Number of observations per state - CPS 2012 dataset

Alabama	680	Kentucky	955	North Dakota	862
Alaska	712	Louisiana	560	Ohio	1504
Arizona	839	Maine	1039	Oklahoma	798
Arkansas	594	Maryland	1824	Oregon	803
California	5866	Massachusetts	971	Pennsylvania	1883
Colorado	1546	Michigan	1349	Rhode Island	1010
Connecticut	1457	Minnesota	1729	South Carolina	765
Delaware	1055	Mississippi	546	South Dakota	1012
District of Columbia	1009	Missouri	971	Tennessee	859
Florida	2630	Montana	519	Texas	3945
Georgia	1414	Nebraska	1207	Utah	827
Hawaii	1183	Nevada	1015	Vermont	949
Idaho	661	New Hampshire	1368	Virginia	1539
Illinois	2115	New Jersey	1376	Washington	1035
Indiana	962	New Mexico	538	West Virginia	590
Iowa	1343	New York	2842	Wisconsin	1259
Kansas	956	North Carolina	1290	Wyoming	924

The 51 clusters in the CPS 2012 dataset correspond to the 50 American states and the District of Columbia. The average number of observations per cluster is 1,288.

List of research reports

15001-EEF: Bao, T., X. Tian, X. Yu, Dictator Game with Indivisibility of Money

15002-GEM: Chen, Q., E. Dietzenbacher, and B. Los, The Effects of Ageing and Urbanization on China's Future Population and Labor Force

15003-EEF: Allers, M., B. van Ommeren, and B. Geertsema, Does intermunicipal cooperation create inefficiency? A comparison of interest rates paid by intermunicipal organizations, amalgamated municipalities and not recently amalgamated municipalities

15004-EEF: Dijkstra, P.T., M.A. Haan, and M. Mulder, Design of Yardstick Competition and Consumer Prices: Experimental Evidence

15005-EEF: Dijkstra, P.T., Price Leadership and Unequal Market Sharing: Collusion in Experimental Markets

15006-EEF: Anufriev, M., T. Bao, A. Sutin, and J. Tuinstra, Fee Structure, Return Chasing and Mutual Fund Choice: An Experiment

15007-EEF: Lamers, M., Depositor Discipline and Bank Failures in Local Markets During the Financial Crisis

15008-EEF: Oosterhaven, J., On de Doubtful Usability of the Inoperability IO Model

15009-GEM: Zhang, L. and D. Bezemer, A Global House of Debt Effect? Mortgages and Post-Crisis Recessions in Fifty Economies

15010-I&O: Hooghiemstra, R., N. Hermes, L. Oxelheim, and T. Randøy, The Impact of Board Internationalization on Earnings Management

15011-EEF: Haan, M.A., and W.H. Siekman, Winning Back the Unfaithful while Exploiting the Loyal: Retention Offers and Heterogeneous Switching Costs

15012-EEF: Haan, M.A., J.L. Moraga-González, and V. Petrikaite, Price and Match-Value Advertising with Directed Consumer Search

15013-EEF: Wiese, R., and S. Eriksen, Do Healthcare Financing Privatisations Curb Total Healthcare Expenditures? Evidence from OECD Countries

15014-EEF: Siekman, W.H., Directed Consumer Search

15015-GEM: Hoorn, A.A.J. van, Organizational Culture in the Financial Sector: Evidence from a Cross-Industry Analysis of Employee Personal Values and Career Success

15016-EEF: Te Bao, and C. Hommes, When Speculators Meet Constructors: Positive and Negative Feedback in Experimental Housing Markets

15017-EEF: Te Bao, and Xiaohua Yu, Memory and Discounting: Theory and Evidence

15018-EEF: Suari-Andreu, E., The Effect of House Price Changes on Household Saving Behaviour: A Theoretical and Empirical Study of the Dutch Case

15019-EEF: Bijlsma, M., J. Boone, and G. Zwart, Community Rating in Health Insurance: Trade-off between Coverage and Selection

15020-EEF: Mulder, M., and B. Scholtens, A Plant-level Analysis of the Spill-over Effects of the German *Energiewende*

15021-GEM: Samarina, A., L. Zhang, and D. Bezemer, Mortgages and Credit Cycle Divergence in Eurozone Economies

16001-GEM: Hoorn, A. van, How Are Migrant Employees Manages? An Integrated Analysis

16002-EEF: Soetevent, A.R., Te Bao, A.L. Schippers, A Commercial Gift for Charity

16003-GEM: Bouwmeerster, M.C., and J. Oosterhaven, Economic Impacts of Natural Gas Flow Disruptions

16004-MARK: Holtrop, N., J.E. Wieringa, M.J. Gijsenberg, and P. Stern, Competitive Reactions to Personal Selling: The Difference between Strategic and Tactical Actions

16005-EEF: Plantinga, A. and B. Scholtens, The Financial Impact of Divestment from Fossil Fuels

16006-GEM: Hoorn, A. van, Trust and Signals in Workplace Organization: Evidence from Job Autonomy Differentials between Immigrant Groups

16007-EEF: Willems, B. and G. Zwart, Regulatory Holidays and Optimal Network Expansion

16008-GEF: Hoorn, A. van, Reliability and Validity of the Happiness Approach to Measuring Preferences

16009-EEF: Hinloopen, J., and A.R. Soetevent, (Non-)Insurance Markets, Loss Size Manipulation and Competition: Experimental Evidence

16010-EEF: Bekker, P.A., A Generalized Dynamic Arbitrage Free Yield Model

16011-EEF: Mierau, J.A., and M. Mink, A Descriptive Model of Banking and Aggregate Demand

16012-EEF: Mulder, M. and B. Willems, Competition in Retail Electricity Markets: An Assessment of Ten Year Dutch Experience

16013-GEM: Rozite, K., D.J. Bezemer, and J.P.A.M. Jacobs, Towards a Financial Cycle for the US, 1873-2014

16014-EEF: Neuteleers, S., M. Mulder, and F. Hindriks, Assessing Fairness of Dynamic Grid Tariffs

16015-EEF: Soetevent, A.R., and T. Bružikas, Risk and Loss Aversion, Price Uncertainty and the Implications for Consumer Search

16016-HRM&OB: Meer, P.H. van der, and R. Wielers, Happiness, Unemployment and Self-esteem

16017-EEF: Mulder, M., and M. Pangan, Influence of Environmental Policy and Market Forces on Coal-fired Power Plants: Evidence on the Dutch Market over 2006-2014

16018-EEF: Zeng,Y., and M. Mulder, Exploring Interaction Effects of Climate Policies: A Model Analysis of the Power Market

16019-EEF: Ma, Yiqun, Demand Response Potential of Electricity End-users Facing Real Time Pricing

16020-GEM: Bezemer, D., and A. Samarina, Debt Shift, Financial Development and Income Inequality in Europe

16021-EEF: Elkhuizen, L, N. Hermes, and J. Jacobs, Financial Development, Financial Liberalization and Social Capital

16022-GEM: Gerritse, M., Does Trade Cause Institutional Change? Evidence from Countries South of the Suez Canal

16023-EEF: Rook, M., and M. Mulder, Implicit Premiums in Renewable-Energy Support Schemes

17001-EEF: Trinks, A., B. Scholtens, M. Mulder, and L. Dam, Divesting Fossil Fuels: The Implications for Investment Portfolios

17002-EEF: Angelini, V., and J.O. Mierau, Late-life Health Effects of Teenage Motherhood

17003-EEF: Jong-A-Pin, R., M. Laméris, and H. Garretsen, Political Preferences of (Un)happy Voters: Evidence Based on New Ideological Measures

17004-EEF: Jiang, X., N. Hermes, and A. Meesters, Financial Liberalization, the Institutional Environment and Bank Efficiency

17005-EEF: Kwaak, C. van der, Financial Fragility and Unconventional Central Bank Lending Operations

17006-EEF: Postelnicu, L. and N. Hermes, The Economic Value of Social Capital

17007-EEF: Ommeren, B.J.F. van, M.A. Allers, and M.H. Vellekoop, Choosing the Optimal Moment to Arrange a Loan

17008-EEF: Bekker, P.A., and K.E. Bouwman, A Unified Approach to Dynamic Mean-Variance Analysis in Discrete and Continuous Time

17009-EEF: Bekker, P.A., Interpretable Parsimonious Arbitrage-free Modeling of the Yield Curve

17010-GEM: Schasfoort, J., A. Godin, D. Bezemer, A. Caiani, and S. Kinsella, Monetary Policy Transmission in a Macroeconomic Agent-Based Model

17011-I&O: Bogt, H. ter, Accountability, Transparency and Control of Outsourced Public Sector Activities

17012-GEM: Bezemer, D., A. Samarina, and L. Zhang, The Shift in Bank Credit Allocation: New Data and New Findings

17013-EEF: Boer, W.I.J. de, R.H. Koning, and J.O. Mierau, Ex-ante and Ex-post Willingness-to-pay for Hosting a Major Cycling Event

17014-OPERA: Laan, N. van der, W. Romeijnders, and M.H. van der Vlerk, Higher-order Total Variation Bounds for Expectations of Periodic Functions and Simple Integer Recourse Approximations

17015-GEM: Oosterhaven, J., Key Sector Analysis: A Note on the Other Side of the Coin

17016-EEF: Romensen, G.J., A.R. Soetevent: Tailored Feedback and Worker Green Behavior: Field Evidence from Bus Drivers

17017-EEF: Trinks, A., G. Ibikunle, M. Mulder, and B. Scholtens, Greenhouse Gas Emissions Intensity and the Cost of Capital

17018-GEM: Qian, X. and A. Steiner, The Reinforcement Effect of International Reserves for Financial Stability

17019-GEM/EEF: Klasing, M.J. and P. Milionis, The International Epidemiological Transition and the Education Gender Gap

2018001-EEF: Keller, J.T., G.H. Kuper, and M. Mulder, Mergers of Gas Markets Areas and Competition amongst Transmission System Operators: Evidence on Booking Behaviour in the German Markets

2018002-EEF: Soetevent, A.R. and S. Adikyan, The Impact of Short-Term Goals on Long-Term Objectives: Evidence from Running Data

2018003-MARK: Gijsenberg, M.J. and P.C. Verhoef, Moving Forward: The Role of Marketing in Fostering Public Transport Usage

2018004-MARK: Gijsenberg, M.J. and V.R. Nijs, Advertising Timing: In-Phase or Out-of-Phase with Competitors?

2018005-EEF: Hulshof, D., C. Jepma, and M. Mulder, Performance of Markets for European Renewable Energy Certificates

2018006-EEF: Fosgaard, T.R., and A.R. Soetevent, Promises Undone: How Committed Pledges Impact Donations to Charity

2018007-EEF: Durán, N. and J.P. Elhorst, A Spatio-temporal-similarity and Common Factor Approach of Individual Housing Prices: The Impact of Many Small Earthquakes in the North of Netherlands

2018008-EEF: Hermes, N., and M. Hudon, Determinants of the Performance of Microfinance Institutions: A Systematic Review

2018009-EEF: Katz, M., and C. van der Kwaak, The Macroeconomic Effectiveness of Bank Bail-ins

2018010-OPERA: Prak, D., R.H. Teunter, M.Z. Babai, A.A. Syntetos, and J.E. Boylan, Forecasting and Inventory Control with Compound Poisson Demand Using Periodic Demand Data

2018011-EEF: Brock, B. de, Converting a Non-trivial Use Case into an SSD: An Exercise

2018012-EEF: Harvey, L.A., J.O. Mierau, and J. Rockey, Inequality in an Equal Society

2018013-OPERA: Romeijnders, W., and N. van der Laan, Inexact cutting planes for two-stage mixed-integer stochastic programs

2018014-EEF: Green, C.P., and S. Homroy, Bringing Connections Onboard: The Value of Political Influence

2018015-OPERA: Laan, N. van der, and W. Romeijnders, Generalized aplha-approximations for two-stage mixed-integer recourse models

2018016-GEM: Rozite, K., Financial and Real Integration between Mexico and the United States

2019001-EEF: Lugalla, I.M., J. Jacobs, and W. Westerman, Drivers of Women Entrepreneurs in Tourism in Tanzania: Capital, Goal Setting and Business Growth

2019002-EEF: Brock, E.O. de, On Incremental and Agile Development of (Information) Systems

2019003-OPERA: Laan, N. van der, R.H. Teunter, W. Romeijnders, and O.A. Kilic, The Data-driven Newsvendor Problem: Achieving On-target Service Levels.

2019004-EEF: Dijk, H., and J. Mierau, Mental Health over the Life Course: Evidence for a U-Shape?

2019005-EEF: Freriks, R.D., and J.O. Mierau, Heterogeneous Effects of School Resources on Child Mental Health Development: Evidence from the Netherlands.

2019006-OPERA: Broek, M.A.J. uit het, R.H. Teunter, B. de Jonge, J. Veldman, Joint Condition-based Maintenance and Condition-based Production Optimization.

2019007-OPERA: Broek, M.A.J. uit het, R.H. Teunter, B. de Jonge, J. Veldman, Joint Condition-based Maintenance and Load-sharing Optimization for Multi-unit Systems with Economic Dependency

2019008-EEF: Keller, J.T. G.H. Kuper, and M. Mulder, Competition under Regulation: Do Regulated Gas Transmission System Operators in Merged Markets Compete on Network Tariffs?

2019009-EEF: Hulshof, D. and M. Mulder, Renewable Energy Use as Environmental CSR Behavior and the Impact on Firm Profit

2019010-EEF: Boot, T., Confidence Regions for Averaging Estimators

2020001-OPERA: Foreest, N.D. van, and J. Wijngaard. On Proportionally Fair Solutions for the Divorced-Parents Problem

2020002-EEF: Niccodemi, G., R. Alessie, V. Angelini, J. Mierau, and T. Wansbeek. Refining Clustered Standard Errors with Few Clusters

www.rug.nl/feb