Explaining Implausible Results in Shadow Economy Estimation Using MIMIC Models

Martina Smrčková[1] | *Prague University of Economics and Business, Prague, Czech Republic*
Karel Brůna[2] | *Prague University of Economics and Business, Prague, Czech Republic*

Received 6.3.2024 (revision received 3.5.2024), **Accepted** (reviewed) 25.6.2024, **Published** 13.9.2024

Abstract

For decades, economists have been trying to estimate the magnitude of the shadow economy (SE), which is not directly observable. This paper explores how the MIMIC (Multiple Indicator Multiple Cause) model can yield estimates of the SE/GDP (the proportion of the SE to the gross domestic product) below 0%, above 100%, and other implausible results. The focus is on the new calibration methods by Dell'Anno (2022) and data on the Czech Republic (1993–2021). The paper concludes that one of the leading causes of implausible results is the misalignment between the SE definition implied by the MIMIC model and that used for the exogenous estimates applied for calibration. Therefore, the authors propose testing the alignment between the SE definitions, such as assessing trends in the latent variable (first-stage scores) resulting from the MIMIC model and the exogenous estimates or applying regression or correlation analysis.

Keywords	***DOI***	***JEL code***
MIMIC model, implausible results, shadow economy, calibration	*https://doi.org/10.54694/stat.2024.12*	*O17, C39, H26*

INTRODUCTION

This paper explains the calibration of shadow economy (SE) estimates derived from the MIMIC (Multiple Indicator Multiple Cause) models, focusing on the implausible results. MIMIC models belong to the most advanced and controversial SE estimation methods. They offer a complex depiction of the economic system, deriving SE estimates as latent variables through the interplay of observable causes and indicators. One of the main reasons for the controversy is that the MIMIC models do not estimate the SE directly. Instead, the latent variable has the form of series of dimensionless indices (first-stage scores) that has to be calibrated to fit an exogenous estimate or series of exogenous estimates.

[1] Department of Monetary Theory and Policy, Faculty of Finance and Accounting, Prague University of Economics and Business, W. Churchill Sq. 4, 120 00 Prague 3, Czech Republic. E-mail: necm02@vse.cz, phone: (+420)732580323. ORCID: <https://orcid.org/0000-0002-0068-5076>.

[2] Department of Monetary Theory and Policy, Faculty of Finance and Accounting, Prague University of Economics and Business, W. Churchill Sq. 4, 120 00 Prague 3, Czech Republic. E-mail: bruna@vse.cz, phone: (+420)224095106. ORCID: <https://orcid.org/0000-0003-3690-1364>.

Implausible calibration results in this paper are defined as calibration results (i.e., the resulting proportion of the SE to the gross domestic product (SE/GDP)) that are economically improbable (such as the resulting SE/GDP being negative or over 100%), do not correspond to expectations based on economic theory and observed phenomena, significantly diverge from exogenous estimates used for calibration, or other estimates that are considered reliable. Finally, results are considered implausible when different calibration methods yield significantly different results.

One of the critical reasons for implausible results may be the misalignment between the SE definition implied by the MIMIC model and the SE definition applied for the exogenous estimates. However, there are other reasons that are connected to economic phenomena, which are briefly discussed in this paper. There is no universally accepted definition of the SE. Instead, the SE is defined by the author of each estimate, whether it is through causes and indicators included in the MIMIC model or formulations of questions in a survey-based estimate.

The misalignment might not necessarily lead to implausible results. However, the reliability of the results may still be dubious since it is unclear which SE definition the resulting SE/GDP represents. Given that the resulting SE/GDP is derived from the first-stage scores based on one SE definition and calibrated to fit exogenous estimates based on another SE definition, drawing any policy-making implications from them or using them for further research is not advisable.

Giles and Tedds (2002) developed the first widely used calibration method, which multiplicatively scales first-stage scores. This method was followed by Dell'Anno and Schneider (2003) and Bajada and Schneider (2005). The main common principle of these three methods (traditional methods) is the use of exactly one exogenous estimate (or its growth rate). Besides that, the methods significantly differ. Dell'Anno and Schneider (2003) shift the time path of the first-stage scores so that the resulting SE/GDP in the base year is equal to the exogenous estimate. In contrast, Bajada and Schneider (2005) add a constant to the first-stage scores (interpreted as growth rates) so that the resulting SE/GDP growth rate equals the exogenous estimate of the SE/GDP growth rate.

Past research (most of which uses the traditional methods) overlooks the need to check the alignment between the SE definition implied by the model and that utilized for the exogenous estimate. While some authors (see Giles and Tedds, 2002; Bajada and Schneider, 2005; and Hassan and Schneider, 2016) use exogenous estimates derived from a separately estimated currency demand model with similar variables to those in the MIMIC model, others (see Dell'Anno, 2007; Nchor, 2021; and Oyibo and Schneider, 2022) use estimates from different methods or variables without documenting the verification of the alignment. Some, see Buehn and Schneider (2008), employed the average of multiple exogenous estimates for the same period. However, this may be an issue when the available estimates are based on different methodologies or (implicit) SE definitions.

In order to battle some deficiencies of the traditional methods, Dell'Anno (2022) developed three new distinct calibration methods. In contrast to the traditional methods, the methods by Dell'Anno (2022) (the new methods) do not merely anchor first-stage scores but adjust estimated coefficients from the MIMIC model using the estimated factor of scale. The new methods also use a more extended time series of exogenous estimates. However, both the traditional and new methods might yield implausible results.

To the authors' knowledge, no research explains and solves the issue of implausible results. Moreover, this issue is rarely discussed in the literature on SE estimation. The only mention of the implausible results the authors were able to find is Feige (2016), who states that Maenhout (2016) attempted to replicate selected estimates from Schneider et al. (2010) and reached negative values of SE/GDP for certain countries based on data for 2006 (–242% for Australia, and –257% for Canada).

The following types of implausible calibration results will be examined and explained in this paper: an inverted trend of the estimated SE/GDP, implausibly low variability of the resulting SE/GDP (i.e., "too flat" time series of the resulting SE/GDP), significant differences between different new calibration methods (especially opposite trends), negative (or implausibly low) and implausibly high values of SE/GDP, and local trends of the resulting SE/GDP that do not correspond to reasonable expectations (e.g., the SE/GDP decreases during a crisis).

This paper aims to give a general framework for recognising the implausible results and explain the reasons for their occurrence from a statistical standpoint. Firstly, it defines and classifies implausible calibration results of the MIMIC model SE estimates. Secondly, it provides answers why implausible results occur in the traditional and new methods. Thirdly, it explains the assumptions the input data must meet for the new methods to yield plausible results and how to use them correctly.

This paper does not cover the economic theories regarding the SE causes and indicators, nor does it extensively discuss the economic phenomena that lead to implausible results. Although it gives several general examples of economic phenomena that may lead to implausible results, it does not explain the specific phenomena in the Czech Republic (1993–2021). However, the authors believe that analysing economic theories and phenomena that lead to implausible results is an important topic that may be the subject of future research.

The first part of this paper explains the structure of the MIMIC model and delineates the reasons for the implausible results observed in both traditional and new calibration methods. The second part briefly describes the MIMIC model estimation for the Czech Republic (1993–2021). Subsequently, the results of the model are calibrated using the new methods and three different types of exogenous estimates: two distinct national accounts-based estimates sourced by Eurostat and the Czech Statistical Office (CZSO), along with a survey-based estimate by Hanousek and Palda (2006).

1 METHODOLOGICAL FRAMEWORK AND CHALLENGES

1.1 Structure of the MIMIC model

The MIMIC model is one of the most advanced methods of the SE estimation since it allows for including multiple SE indicators and not just one – such as the currency demand or electricity consumption approaches - and exploits the economic theory on causal relationships between SE and its observable causes and indicators.

The MIMIC model consists of a structural model linking the latent variable to its causes and a measurement model linking the latent variable to its indicators. The structural model in Formula (1) describes the relationship between the latent variable and its causes:

$$SE_t = \sum_{i=1}^{C} \gamma_i X_{i,t} + \zeta_t \;, \tag{1}$$

where SE_t is the SE first-stage score at time t, γ_i are the structural coefficients, $X_{i,t}$ is the value of the cause i in time t for $i \in \{1, 2, \ldots, C\}$, ζ_t is the error component at time t and C is the total number of causes.

The measurement model is the equation system that links the indicators to the latent variable. The MIMIC model also allows for the addition of direct relationships between some causes and indicators, where the cause explains the indicator. For example, a model may have four causes: tax burden, unemployment rate, self-employment, and economic freedom, and two indicators: GDP, and cash outside of banks. The unemployment rate may directly influence the GDP, and not just through the SE. Formula (3) is an example of a measurement equation with a direct relationship.

$$Y_{1,t} = \lambda_1 SE_t + \varepsilon_{1,t} \;, \tag{2}$$

$$Y_{2,t} = \lambda_2 SE_t + \delta_1 X_{1,t} + \varepsilon_{2,t} \;, \tag{3}$$

$$\ldots$$

$$Y_{P,t} = \lambda_P SE_t + \varepsilon_{P,t} \;, \tag{4}$$

where $Y_{p,t}$ are values of the indicators at time t, P is the total number of indicators, λ_p are the measurement coefficients, δ_1 is the coefficient describing the direct relationship between the indicator Y_2 and the cause X_1, and $\varepsilon_{p,t}$ are the error components at time t for $p \in \{1, 2, \ldots, P\}$.

The MIMIC specification assumes that the error component in the structural equation (Formula 1) ζ_t is not correlated with causes, the error components in the measurement model $\varepsilon_{p,t}$ do not correlate either with causes or with the latent variable, and the error component in the structural equation ζ_t does not correlate with the error components in the measurement equations ε_p (Buehn and Schneider 2008).

The parameters of the MIMIC model are not just the structural and measurement coefficients (including intercepts) but also the latent variable (SE) variance, residual variances from the structural and measurement equations, and covariances between the causes. Unless either one of the measurement coefficients or the latent variable's variance is constrained,[3] the MIMIC procedure yields infinite solutions to the structural and measurement coefficients ($\gamma_1, \ldots, \gamma_C$ and $\lambda_1, \ldots, \lambda_P$).

The variance of the latent variable is usually constrained to 1 and the measurement coefficient to ± 1. The indicator whose coefficient is constrained is called the reference indicator, and it is usually the indicator that is expected to have the highest correlation with the SE. The latent variable, therefore, has the same unit of measure as the reference indicator. The sign of the constrained measurement coefficient is chosen based on the expected relationship between the SE and the reference indicator. This paper constrains one of the measurement coefficients since it is more common.

The MIMIC model is a model specification of a structural equation model (SEM).[4] In this paper, the estimation is done using the covariance-based approach, which minimises the distance between the sample covariance matrix that contains variances and covariances between the observed causes and indicators, and the covariance matrix predicted by the model. There are different algorithms to estimate the coefficients. In this paper, the maximum likelihood (ML) function is used.

1.2 The traditional methods: methodology and causes for implausible results

Calibration converts first-stage scores from the MIMIC model to SE/GDP values. The MIMIC models capture the economic phenomena influencing the SE. In contrast, calibration is a purely mathematical exercise that aims to make the MIMIC model output more economically interpretable. The calibration methods are not directly tied to economic theories but entail assumptions, elaborated later. Failing these may yield implausible results. Possible causes of not meeting the assumptions are changes in the economy's structure during the examined period (such as tax reforms or laws reducing the SE).

Firstly, the first-stage scores are calculated as:

$$\hat{SE}_t^{FS\_est} = \sum_{i=1}^{C} \hat{\gamma}_i X_{i,t}, \tag{5}$$

where $\hat{SE}_t^{FS\_est}$ is the first-stage latent score at time t, $\hat{\gamma}_i$ are the ML estimates of the structural coefficients from equation (1), $X_{i,t}$ is the value of the cause i in time t, and C is the total number of causes used in the final model. Giles and Tedds (2002) used standardised data for the estimation. However, they calculated the first-stage scores with raw data, while Dell'Anno and Schneider (2003), and Bajada and Schneider (2005) used deviations from means of the variables.

Traditional calibration methods use only one or two exogenous estimates. The exogenous estimate is a SE/GDP estimate obtained separately from the model. It may be, for example, an estimate from an official authority (such as the estimates by Eurostat or CZSO used in this paper), a survey

3 Based o the authors' knowledge and the current literature, these are the only two types of parameters whose constraint can solve the identification issue.

4 For more information on SEMs, see, for example, Bollen (1989).

(e.g., the estimate by Hanousek and Palda (2006) used in this paper), or a model-based estimate published. The fundamental principle of the traditional calibration methods is that the estimated SE/GDP, or its growth rate in the calibration method by Bajada and Schneider (2005), in the base period (i.e., the period for which the exogenous estimate is available) must be equal to the exogenous estimate.

The most widely used calibration method is the one by Giles and Tedds (2002), which multiplicatively scales the first-stage scores so that the resulting SE/GDP in the base year equals the exogenous estimate, as seen in Formula (6). Therefore, this method assumes a constant ratio of the SE to the first-stage scores. The SE/GDP is:

$$\hat{SE}_t = \frac{SE_{t*}^{exog}}{\hat{SE}_{t*}^{FS\_est}} \hat{SE}_t^{FS\_est}, \tag{6}$$

where $\hat{SE}_t$ is the resulting SE/GDP at time t, SE_{t*}^{exog} is the exogenous estimate, t^* is the base period, and $\hat{SE}_t^{FS\_est}$ is the first-stage score at time t.

The exogenous estimate's level determines the resulting SE/GDP's level, while the first-stage scores determine the time path. However, the standard deviation of the resulting SE/GDP is not equal to the standard deviation of the first-stage scores but is influenced by the $\frac{SE_{t*}^{exog}}{\hat{SE}_{t*}^{FS\_est}}$. The higher the SE_{t*}^{exog} than the $\hat{SE}_{t*}^{FS\_est}$, the higher the standard deviation of the resulting SE/GDP. Therefore, the resulting SE/GDP's implausibly low variability ("a too flat time series") occurs when the first-stage score in the base period is significantly higher than the exogenous estimate.

On the other hand, if the exogenous estimate is much higher than the first-stage score in the base period, the slope of the estimated SE/GDP may be very steep, and implausibly low or high values of the SE/GDP may occur. An inverted trend occurs when the first-stage score is negative in the base period, which might happen depending on the estimated coefficients or the values of the causes.

The method by Dell'Anno and Schneider (2003) calibrates the first-stage scores by adding a constant, which preserves absolute differences between the first-stage scores. Therefore, this method assumes a constant difference between the the SE and the corresponding first-stage score. This method can be applied only when the model is estimated in differences (i.e., the first-stage scores are interpreted as the first difference of the SE/GDP). The resulting SE/GDP is calculated as follows:

$$\hat{SE}_{t*} = SE_{t*}^{exog}, \tag{7}$$

$$\hat{SE}_t = \hat{SE}_{t-1} + \sum_{i=1}^{C} \hat{\gamma}_i \Delta X_{i,t} \text{ for } t > t_*, \tag{8}$$

$$\hat{SE}_t = \hat{SE}_{t+1} - \sum_{i=1}^{C} \hat{\gamma}_i \Delta X_{i,t+1} \text{ for } t < t_*. \tag{9}$$

The implausibly low variability and inverted trend are impossible since the first-stage scores are shifted, so the SE/GDP in the base period equals the exogenous estimate. However, implausibly high or low values of the resulting SE/GDP may be an issue depending on the difference between the first-stage score in the base period and the exogenous estimate.

The method by Bajada and Schneider (2005) calibrates the growth rate of the SE/GDP so that the growth rate in the base period is equal to the growth rate of the exogenous estimates. The model is specified and estimated in terms of growth rates. This method needs two consecutive exogenous estimates to calculate the growth rate. The constant $\tilde{g}_{t*}$ is calculated as:

$$\tilde{g}_{t*} = g_{t*}^{exog} - g_{t*}^{FS\_est}, \tag{10}$$

where g_{t*}^{exog} is the exogenous growth rate, and $g_{t*}^{FS\_est}$ is the base year's first-stage score growth rate. Then, the constant $\tilde{g}_{t*}$ is added to the growth rates of the first-stage scores. Therefore, this method assumes a constant difference between the growth rates of the first-stage scores and the SE. The SE/GDP growth rates are calculated as:

$$\hat{g}_t = \tilde{g}_{t*} + g_t^{FS\_est}, \tag{11}$$

where $\hat{g}_t$ is the growth rate of the estimated SE/GDP at time t. The growth rates are then employed to get the estimates of SE/GDP levels.

The inverted trend (i.e., the estimated SE/GDP growth rates having opposite signs to first-stage scores) can occur when the growth rate of the first-stage scores in the base period and the growth rate of the exogenous estimates have opposite signs. However, this does not mean all negative growth rates change into positives. If the first-stage score at the time t is less than the constant $\tilde{g}_{t*}$, the resulting SE/GDP at time t can still be negative. The opposite signs (as well as vastly different growth rates of the first-stage scores and exogenous estimates) indicate a misalignment between the SE definitions implied by the MIMIC model and the exogenous estimates.

The resulting SE/GDP variability heavily depends on the constant $\tilde{g}_{t*}$ since it accumulates exponentially. If the constant $\tilde{g}_{t*}$ is high, the estimated SE/GDP rises exponentially, and some values of the resulting SE/GDP may be implausibly high. Conversely, the variability may be implausibly low if the constant is negative.

The misalignment between the SE definition implied by the MIMIC model and the SE definition used for the exogenous estimate may cause issues even when applying the traditional methods. Suppose the exogenous estimate (or its growth rate) is available for more periods. In that case, the authors advise testing the alignment by comparing the trends, correlation, or regression with the exogenous estimate as the explained variable and the first-stage scores as the explanatory variable. If the tests show misalignment, it is not advisable to perform the calibration. Instead, the authors advise using a different exogenous estimate corresponding to the SE definition implied by the model or redeveloping the MIMIC model so that the causes and indicators correspond more to the SE definition used for the exogenous estimate. The authors recommend testing for robustness by calibrating with multiple base years if possible and seeing if the resulting SE/GDP levels and time paths differ significantly.

Suppose there is only one exogenous estimate available. In that case, the authors recommend carefully examining the methodology used for the exogenous estimate and checking if the causes and indicators in the MIMIC model reflect the same SE definition.

1.3 The new methods: detailed methodology and data issues leading to implausible results

The new calibration methods by Dell'Anno (2022) provide several substantial advantages over the traditional methods. Firstly, they do not merely anchor first-stage scores but adjust the estimated coefficients using an estimated scale factor. Secondly, they use a more extended time series of exogenous estimates with at least two observations. Thirdly, the results of these methods are not sensitive to the level of first-stage scores (i.e., the resulting SE/GDP is the same when a constant is added to the first-stage scores). Fourthly, they allow for thorough testing of the alignment of the SE definition implied by the MIMIC model with the definition used for the exogenous estimates.

Dell'Anno (2022) developed three calibration methods. However, Method 3 gives the same results as Method 1. Therefore, it will not be discussed in this paper. Since Dell'Anno (2022) developed the methods for raw data and this paper works with standardised data, the first-stage scores are calculated as follows. Formula (12) then replaces Formula (5):

$$\hat{SE}_t^{FS\_est} = \sum_{i=1}^{C} \hat{\gamma}_i \frac{\sigma_{Y_1}}{\sigma_{X_i}} X_{i,t} , \tag{12}$$

where $\sigma_{Y_{1,t}}$ is the standard deviation of the reference indicator, σ_{X_i} is the standard deviation of the i-th cause, $X_{i,t}$ are the non-standardised values of the causes.

Method 1 minimises the difference between the first-stage scores and exogenous estimates. Firstly, the exogenous estimates are regressed on the first-stage latent score:

$$SE_{t*}^{exog} = \rho_0 + \rho_1 \hat{SE}_{t*}^{FS\_est} + a_{t*} \text{ with } t^* \in W , \tag{13}$$

where SE_{t*}^{exog} are the exogenous estimates of SE/GDP, ρ_0 and ρ_1 are auxiliary regression coefficients, a_{t*} is the error component, $W = \{w_1, \ldots, w_e\}$ are years for which exogenous estimates are available, and e is the total number of exogenous estimates. This method can only be used if it is possible to obtain $\hat{\rho}_0$ and $\hat{\rho}_1$, OLS (ordinary least squares) estimates of ρ_0 and ρ_1. Then $\hat{\rho}_1$, is used to rescale the structural coefficients as:

$$^{est\_1}\hat{\gamma}_i^* = \hat{\gamma}_i \hat{\rho}_1 , \tag{14}$$

where $^{est\_1}\hat{\gamma}_i^*$ is the rescaled coefficient for the i-th cause for Method 1. This method estimates the regression coefficients for years with available exogenous estimates and then uses them to alter the first-stage scores in the entire period. Therefore, this method assumes a constant linear relationship between the first-stage scores and the SE (meaning constant variability of the first-stage scores and constant correlation between the first-stage scores and the exogenous estimates).

The final step is to calculate the resulting SE/GDP using the estimated intercept from Formula (13) and rescaled coefficients as in the equations below:

$$^{est\_1}\hat{SE}_t = \hat{\rho}_0 + \sum_{i=1}^{C} {}^{est\_1}\hat{\gamma}_i^* \frac{\sigma_{Y_1}}{\sigma_{X_i}} X_{i,t} , \tag{15}$$

where $^{est\_1}\hat{SE}_t$ is the resulting SE/GDP from Method 1 at time t.

Method 2 minimises the difference between the means of the first-stage scores and exogenous estimates. This time, the reference indicator is regressed on the exogenous estimates:

$$Y_{1,t} = const + {}^{REG}\lambda_1 SE_{t*}^{exog} + e_{t*} \text{ with } t^* \in W , \tag{16}$$

where $Y_{1,t*}$ is the reference indicator, const is the intercept, $^{REG}\lambda_1$ is the "coefficient of scale", the true value of the reference coefficient (in the MIMIC model, the reference coefficient was constrained to 1), and e_{t*} is the error component. This method can only be used if it is possible to obtain $\widehat{const}$ and $^{REG}\hat{\lambda}_1$, OLS estimates of *const* and $^{REG}\lambda_1$.

Then, the structural coefficients are rescaled as follows:

$$ {}^{est\_2}\hat{\gamma}_i^* = \frac{\hat{\gamma}_i}{{}^{REG}\hat{\lambda}_1} , \tag{17} $$

where ${}^{est\_2}\hat{\gamma}_i^*$ is the rescaled coefficient for the *i*-th cause for Method 2.

This method assumes a constant linear relationship between the reference indicator and the SE (constant variability of the reference indicator and the first-stage scores and constant correlation between the reference indicator and the exogenous estimates).

The next step is to estimate the intercept that is later added to the rescaled first-stage scores as:

$$ {}^{\Delta\mu\_est}\hat{\gamma}_0^* = Mean(SE_W^{exog}) - Mean\left(\frac{\hat{SE}_W^{FS\_est}}{{}^{REG}\hat{\lambda}_1}\right) , \tag{18} $$

where ${}^{\Delta\mu\_est}\hat{\gamma}_0^*$ is the intercept, $Mean(SE_W^{exog})$ is the mean of exogenous estimates, $Mean(\hat{SE}_W^{FS\_est})$ is the mean of the first-stage scores in periods with exogenous SE/GDP estimates available.

The resulting SE/GDP is described as:

$$ {}^{est\_2}\hat{SE}_t = {}^{\Delta\mu\_est}\hat{\gamma}_0 + \sum_{i=1}^{C} {}^{est\_2}\hat{\gamma}_i^* \frac{\sigma_{Y_1}}{\sigma_{X_i}} X_{i,t} , \tag{19} $$

where ${}^{est\_2}\hat{SE}_t$ is the resulting SE/GDP from Method 2 at time *t*.

The inverted trend refers to a scenario where the trend resulting from the calibration is opposite to that of the first-stage scores. This occurs when the OLS estimate of the auxiliary regression coefficient (ρ_1 in Formula (13) for Method 1 or ${}^{REG}\lambda_1$ in Formula (16) for Method 2) is negative. Rescaling structural coefficients reverses the signs that no longer make economic sense. Notably, observations with available exogenous SE estimates determine the regression coefficients. Even though the global trends of the exogenous estimates and the first-stage scores (or the reference indicator) are accordant, the local trends for the observations with exogenous SE estimates available might differ, turning the regression coefficient negative.

Implausibly low variability of the estimated SE/GDP means that the estimated SE/GDP (from Formula (15) or (19)) has a standard deviation significantly lower than the first-stage scores. The authors have decomposed the standard deviation of the estimated SE/GDP (for the derivation of Formula (20) and (21) see Annex B). For Method 1, the standard deviation of SE/GDP is:

$$ \sigma_{{}^{est\_1}\hat{SE}} = \sigma_{SE_W^{exog}} \frac{\sigma_{\hat{SE}^{FS\_est}}}{\sigma_{\hat{SE}_W^{FS\_est}}} \left| corr(\hat{SE}_W^{FS\_est}, SE_W^{exog}) \right| , \tag{20} $$

where $\sigma_{{}^{est\_1}\hat{SE}}$ is the standard deviation of the resulting SE/GDP according to Method 1, $\sigma_{SE_W^{exog}}$ is the standard deviation of the exogenous estimates, $\sigma_{\hat{SE}^{FS\_est}}$ is the standard deviation of the first-stage scores, $corr(\hat{SE}_W^{FS\_est}, SE_W^{exog})$ is the correlation between first-stage scores and the exogenous estimates.

Therefore, implausibly low variability of the estimated SE/GDP may be caused by the following factors (or their combination): low correlation between the first-stage scores and the exogenous estimates, a low standard deviation of the exogenous estimates, or the standard deviation of the first-stage scores for observations with exogenous SE estimates available being significantly higher than the standard deviation of the first-stage scores for the whole period. If the standard deviation of the first-stage scores is invariant throughout the entire period, it does not influence the standard deviation of the result suggested by Method 1.

For Method 2, the standard deviation of SE/GDP is:

$$\sigma_{est\_2_{\hat{SE}}} = \sigma_{SE_W^{exog}} \frac{\hat{\sigma}_{\hat{SE}^{FS\_est}}}{\sigma_{Y_{1,W}}} \frac{1}{|\,corr(Y_{1,W}, SE_W^{exog})|}, \tag{21}$$

where $\sigma_{est\_2_{\hat{SE}}}$ is the standard deviation of the resulting SE/GDP according to Method 2, $\sigma_{Y_{1,W}}$ is the standard deviation of the reference indicator, and $corr(Y_{1,W}, SE_W^{exog})$ is the correlation between the reference indicator and the exogenous estimates.

Therefore, implausibly low variability of the estimated SE/GDP may be caused by the following factors (or their combination): high correlation between the first-stage scores and the exogenous estimates (only to a certain extent), a low standard deviation of the exogenous estimates, or the standard deviation of the reference indicator for observations with exogenous SE estimates available being significantly higher than the standard deviation of the first-stage scores.

There are two types of significant differences between the results of the two methods: significantly different variabilities, and opposite trends. The resulting SE/GDP of both methods are equal when (for the derivation, see Annex C):

$$\frac{1}{\sigma_{\hat{SE}_W^{FS\_est}}} |corr(\hat{SE}_W^{FS\_est}, SE_W^{exog})| = \frac{1}{\sigma_{Y_{1,W}}} \frac{1}{|corr(Y_{1,W}, SE_W^{exog})|}. \tag{22}$$

In the ideal case, the exogenous estimates' correlation with both the first-stage scores and the reference indicator is 1, and the standard deviation of the first-stage scores is equal to the standard deviation of the reference indicator. Yet, the equation can hold even if $\frac{1}{\sigma_{\hat{SE}_W^{FS\_est}}}$ is A times greater than $\frac{1}{\sigma_{Y_{1,W}}}$, and simultaneously $\frac{1}{|corr(Y_{1,W}, SE_W^{exog})|}$ is A times greater than $|corr(\hat{SE}_W^{FS\_est}, SE_W^{exog})|$.

The opposite trends of the resulting SE/GDP from Method 1 and Method 2 occur when there is an inverted trend in one method but not in the other. This is when exogenous estimates positively correlate with first-stage scores (the reference indicator) and negatively with the reference indicator (first-stage scores).

Implausibly low (high) values of SE/GDP may occur when the SE/GDP has a steep upward trend and there are low (high) exogenous values towards the end (beginning) of the time series. Alternatively, the same may occur when the SE/GDP has a steep downward trend and there are low (high) exogenous values towards the beginning (end) of the time series. The slope's steepness is determined by the resulting SE/GDP variability.

Local trends of the SE/GDP not corresponding to reasonable expectations (e.g., the SE/GDP is expected to rise during a crisis) may be caused by the characteristics of the exogenous estimates, characteristics of the first-stage scores, or interactions between the two. The local trend of the exogenous estimates itself may not correspond to expectations, which could result from the methodology and the SE definition implied by the exogenous estimates. If the cause are the first-stage scores, it may point to an incorrect model specification or an implied SE definition, according to which the local trend does not correspond to expectations. The local trend not corresponding to expectations may also be caused by the inverted trend.

1.4 Potential causes for the implausible results of the new methods

The auxiliary regressions in the new methods can be interpreted as testing the alignment of the SE definition implied by the model to the SE definition used for the exogenous estimates. The authors consider a good regression fit when R^2 is at least 0.6 and the estimated coefficient $\hat{\rho}_1$ for Method 1 or ${}^{REG}_{\square}\hat{\lambda}_1$ for Method 2 is significant at least at the 10% level. Suppose these conditions are not met. In that case, the first-stage scores do not explain the exogenous estimates well (Method 1), or the exogenous estimates do not explain the reference indicator well (Method 2).

Both could be signs that the SE definition implied by the model (either through the causes or the reference indicator) is not aligned with the SE definition used for the exogenous estimates. Some of the causes of implausible results in these methods can be traced to the poor fit of the auxiliary regressions.

However, implausible results could also result from structural economic breaks that happened after the period of the last exogenous estimate or before the period of the first exogenous estimate and change the relationship between the SE and the first-stage scores or the reference indicator. For example, a new policy makes involvement in the SE more difficult. In this case, the SE may decrease without a decrease in the first-stage scores or in the reference indicator.

Another issue are influential observations in the auxiliary regressions. An influential observation does not have to be influential from the economic standpoint (i.e., there was a significant event). There may be the following reasons for the occurrence of influential points.

Firstly, the SE/GDP may have been incorrectly measured in a particular period. Secondly, the methodology for the exogenous estimates may have changed. Thirdly, an extreme value of one or more causes may have occurred that did not influence the SE. Fourthly, the exogenous estimates capture something that the first-stage scores do not, and vice versa. One explanation for this may be a misalignment in the SE definitions. Another explanation may be a phenomenon captured by the model and not by the exogenous estimates but not directly connected to the SE definitions.

This may happen, for example, when increased unemployment in one sector leads to increased involvement in the SE for workforce that has formerly worked in that sector. The model captured this shock through the unemployment rate. However, the exogenous estimates were based on a survey of respondents who work in different sectors.

Another example may be that the exogenous estimates accurately captured an SE increase, but the model did not. For example, the model has three SE causes: tax burden, unemployment rate, and regulatory burden. However, the SE increase may have been caused by an immigration wave, resulting in many immigrants starting to work informally, at least initially. In this case, the first-stage scores would increase, unlike the exogenous estimates.

These phenomena that cause influential points in the auxiliary regressions may also cause an issue when applying traditional methods when the exogenous estimate is available for a period when there is an influential point either in the exogenous estimate or the first-stage scores. However, this paper will not further examine economic phenomena that may lead to implausible results. This paper primarily explores the statistical patterns that lead to implausible results and tries to derive a general framework for classifying and discovering them. Nevertheless, economic phenomena leading to implausible results are an important topic that may be the subject of future research.

2 MODEL BUILDING FOR CZECH CONDITIONS

This section briefly explains the MIMIC model estimation for the Czech Republic from 1993 to 2021. All the variables considered for the analysis have economic reasoning behind them. Nonetheless, this paper does not aim to explore economic theories regarding the SE causes and indicators. Instead, it aims to explain implausible results in the SE estimation using the MIMIC model, which is used only as a tool.

The authors divided causes into seven segments: taxation, job market, freedom, education, the economic situation of businesses, the economic situation of households, and others. The complete list of causes is in Table A1 in the Annex A. The list of causes and indicators used in the reported models is in Table 1, along with details on the data sources and adjustments.

Three types of indicators were used: ratio of value added (VA) per hour worked in the four sectors with the highest SE relative to VA per hour worked in the four sectors with the lowest SE, real household consumption per capita, and a monetary variable (either Cash outside of banks / M1 or Card payments value / M1). Since the first indicator is newly used in this paper, it is explained in more detail in this section than the other variables.

The four sectors with the highest and lowest SE were determined from the SE estimation by sector for the Czech Republic for 2018 from the CIRCABC database (Eurostat, 2023). The sectors with the highest SE are: T – Activities of households as employers, A – Agriculture, forestry and fishing, F – Construction, and I – Accommodation and food service activities. The four sectors with the lowest SE are: B – Mining and quarrying, K – Financial and insurance activities, P – Education, and Q – Human health and social work activities.

VA per hour worked in the four sectors with the highest SE relative to VA per hour worked in the four sectors with the lowest SE approximates undeclared work. In order to calculate VA per hour worked, the VA is divided only by the declared hours. Therefore, the declared VA per hour worked increases even though the same VA is produced with the same number of hours just because of increased hours worked informally (i.e., the VA was divided by a lower number of declared hours). The VA per hour worked of the four sectors with the highest SE is divided by the VA per hour worked in the four sectors with the lowest SE to control for technological progress and changes in labour productivity. This paper assumes that the four sectors with the highest SE are influenced by technological progress, and their labour productivity grows at the same rate to those with the lowest SE. The authors are aware of this indicator's deficiencies. However, they believe it still partially captures the SE as one of the three indicators in the model.

The authors also tried two, three, and five sectors with the highest and lowest SE. However, the variant with four sectors leads to the best fit of the models. It is also noteworthy that including too few sectors may make the indicator unstable and influenced too much by the development of individual sectors.

Household consumption per capita was preferred over GDP in this study due to its closer alignment with the SE, as GDP encompasses the non-SE-related components such as government procurement and international trade. The authors used the Card payments value / M1 as an alternative to Cash outside of banks / M1, supported by recent findings such as those of Marmora and Mason (2021), indicating a significant negative association between SE and electronic payments.

Table 2 contains the best model results and a model used for robustness check. Many other models were estimated but discarded due to one of the following reasons: model not converging, negative residual variance, rejected hypothesis of the chi-square test, RMSEA (Root Mean Square Error of Approximation) greater than 0.1, not economically justifiable signs of the coefficients, Akaike information criterion (AIC) or Bayesian information criterion (BIC) higher than the best model's.

Furthermore, three more modifications of the best model were estimated for robustness check. The modifications are changing the Fraser Institute Freedom index for the economic freedom index from the Heritage Foundation, changing the NPL ratio of loans to households for the NPL ratio of consumer loans, and changing the Card payments value / M1 for Cash outside of banks / M1. Only the results for the last modification are shown here because the changes in the model and calibration results are most apparent for that case.

While the coefficient at the Card payments value / M1 is negative, the coefficient at Cash outside of banks / M1 is positive. All the other coefficients for causes and indicators have the same signs in both models. However, in the robustness check, the coefficient for the lagged unemployment rate is insignificant. Moreover, the model used for the robustness check has an RMSEA slightly greater than 0.1, which is not considered a good fit according to the rule of thumb by Browne and Cudeck (1993). However, the hypothesis of the chi-square test is not rejected.

Table 1 Data description of causes and indicators used in the best model and the model used for the robustness check

Variable name	Data source	Group	Range of data available	Data adjustment
Causes				
Fraser Institute Freedom index	Fraser Institute database	Freedom	1995, 2000–2020	Linear interpolation, extrapolation by taking the last known difference between observations
Unemployment rate (%)	Czech Statistical Office statistics	Job market	1993–2021	–
Public social spending / GDP (%)	OECD database	Others	1993–2021	–
NPL ratio of loans to households (%)	Czech National Bank – the ARAD time series database	Economic situation of households	2002–2021	Linear extrapolation
Indicators				
VA per hour worked in the four sectors with the highest SE relative to VA per hour worked in the four sectors with the lowest SE	Czech Statistical Office statistics	NA	Values added: 1993–2021 Hours worked: 1995–2021	Hours worked: extrapolation by taking the last known difference between observations
Household consumption per capita	Czech Statistical Office statistics	NA	1993–2021	–
Card payments value / M1 (%)	Card payments value: European Central Bank data portal	NA	Cards payments value: 2000–2021 M1: 1993–2021	Cards payments value: exponential extrapolation
Cash outside of banks / M1 (%)	M1: Federal Reserve economic data	NA	Cash outside of banks: 2002–2021 M1: 1993–2021	Cash outside of banks: linear extrapolation

Notes: NA means not applicable. The indicators were not divided into groups.
Source: Own construction

Table 2 Selected MIMIC estimation results

	The best model	Robustness check
Causes		
Fraser Institute Freedom index	–0.19***	–0.29***
	(–3.05)	(–2.61)
Unemployment rate (%) (t – 1)	0.21***	0.02
	(3.2)	(0.85)
Public social spending/GDP (%) (t – 1)	–0.25***	–0.12**
	(–3.15)	(–2.34)
NPL ratio of loans to households (t – 1)	0.07***	0.08**
	(2.74)	(2.35)
Indicators		
VA per hour worked in the four sectors with the highest SE relative to VA per hour worked in the four sectors with the lowest SE	1.00	1.00
	(NA)	(NA)

Table 2 (continuation)

	The best model	Robustness check
Indicators		
Card payments value/M1 (%)	−1.91***	–
	(−3.3)	–
Real final consumption expenditure of households per capita	−3.35***	−2.3***
	(−3.13)	(−2.62)
Cash outside of banks/M1 (%)	–	2.09***
	–	(2.67)
Direct relationships		
Real final consumption expenditure of households per capita ~ Unemployment rate (%) (t – 1)	0.51***	−0.18***
	(5.11)	(−4.06)
Real final consumption expenditure of households per capita ~ Public social spending/GDP (%) (t – 1)	−0.67***	−0.11
	(−4.07)	(−1.26)
Fit measures		
Converged	YES	YES
Negative variance	NO	NO
Chi-square (model vs. saturated)	7.98	12.43
DF	9.00	9.00
p-value	0.54	0.19
Chi-square (baseline vs. saturated)	234.42	225.63
DF	15.00	15.00
p-value	0.00	0.00
RMSEA	0.00	0.12
AIC	27.07	36.63
BIC	43.06	52.61

Notes: The asterisks denote p-values of the estimated coefficients, *** p<0.01, ** p<0.05, * p<0.1. Z-scores are in parentheses. Since standardised data was used, all intercepts are zero and not reported.
Source: Own construction

3 DEMONSTRATION OF THE IMPLAUSIBLE RESULTS

3.1 Exogenous estimates

This paper uses three different exogenous estimates: a national accounts-based estimate by Eurostat (2005 and 2023), a national accounts-based estimate from the CZSO (Czech Statistical Office, 2023), and a survey-based estimate from Hanousek and Palda (2006). The objective was to use exogenous estimates that are not model-based. Some authors believe (see Kirchgässner, 2016; Dell'Anno, 2022) that national accounts or survey-based estimates are more reliable than econometric methods (such as the currency demand method and MIMIC models). However, estimates from national statistical offices and surveys are often available only for selected periods and might be published with a significant time lag. On the other hand, the MIMIC models can predict the SE/GDP for periods for which estimates from surveys or national statistical offices are not available.

The authors do not advise using the MIMIC estimates as exogenous estimates since these estimates are based on another exogenous estimate. In the authors' opinion, using an original exogenous estimate is always better than an estimate derived from it. A crucial weakness of the currency demand method is that it captures only the one-sided relationship between the SE and the amount of currency and does not consider SE components that are uncorrelated to using cash.

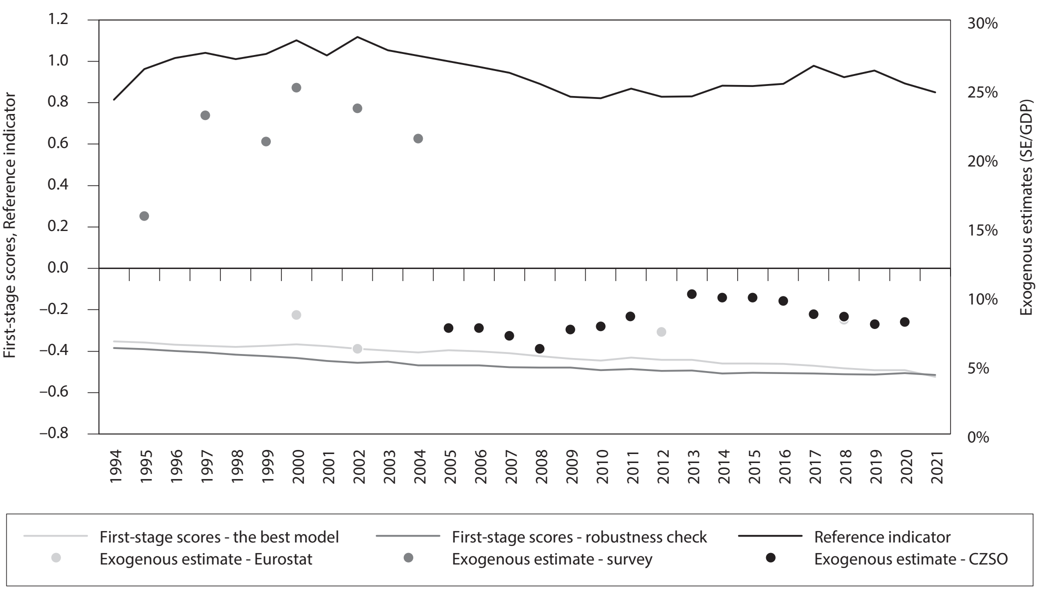
The national accounts-based method by Eurostat is based on adjusting the estimated GNP by non-exhaustiveness. There are the following seven types of non-exhaustiveness: N1 – Underground production, N2 – Illegal production, N3 – Informal production, N4 – Production of households for own final use, N5 – Statistically underground production, N6 – Producers deliberately misreporting, N7 – Deficiencies in the statistical system. The types of non-exhaustiveness are separately estimated using the labour input method, commodity flow method, special surveys, fiscal and other audit data, and other methods. Eurostat estimates are available for 2000, 2002, 2012, and 2018.

CZSO uses a similar national accounts-based method as Eurostat with slight differences in types of non-exhaustiveness; N4 is not part of the estimate, and N7 is defined more specifically as wages and salaries in kind (which are also part of the N7 as defined by Eurostat). For this paper, N2 and N7 were excluded by Eurostat and CZSO estimates since they are outside the scope of the SE according to most definitions.

Hanousek and Palda's (2006) estimate is based on a survey carried out in 2000, 2002, and 2004, wherein respondents were asked about their SE participation with response options categorized as never, sometimes, or often. The latter two responses were considered indicative of the SE involvement, while the former indicated non-participation. The respondents were asked about their informal activities in the current year and two and five years prior. Therefore, the estimates are available for 1995, 1997, 2000, 2002, and 2004. Two estimates were available for each of the years 1999, 2000, and 2002, the average of which was used.

Figure 1 shows the three exogenous estimates, the first-stage scores from models presented in Table 2, and the reference indicator. Table 3 shows correlations between the exogenous estimates, the first-stage scores, and the reference indicator. While the Eurostat and CZSO estimates are very close

Figure 1 Exogenous estimates, first-stage scores, and the reference indicator



Note: Sources of the exogenous estimates: Eurostat (2005), Eurostat (2023), Hanousek and Palda (2006), Czech Statistical Office (2023).
Source: Own construction

and range between 6 and 10%, the survey-based estimates are much higher (between 15 and 25%). The considerable differences in levels are due to different methodologies and likely due to different implied SE definitions.

Furthermore, Figure 1 shows that the survey-based estimate has a similar time path to the reference indicator, which means that they are probably based on comparable SE definitions. Moreover, Table 3 shows that they are highly correlated. In contrast, the Eurostat and CZSO estimates do not show the same trend as the reference indicator. The first-stage scores from both models have similar time paths and are highly correlated. However, the first-stage scores do not follow the same trend as any of the exogenous estimates and are negatively correlated with all of them.

Table 3 Correlations between the exogenous estimates, first-stage scores from the final model and the model used for robustness checking, and the reference indicator

	Exogenous estimates – Eurostat	Exogenous estimates – survey	Exogenous estimates – CZSO	First-stage scores – the best model	First-stage scores – robustness check	Reference indicator
Exogenous estimates – Eurostat	1.000					
Exogenous estimates – survey	NA	1.000				
Exogenous estimates – CZSO	NA	NA	1.000			
First-stage scores – the best model	–0.165	–0.299	–0.424	1.000		
First-stage scores – robustness check	–0.029	–0.517	–0.599	0.915	1.000	
Reference indicator	–0.166	0.909	–0.238	0.543	0.455	1.000

Notes: NA means that the correlation could not be calculated due to a lack of overlapping observations.
Source: Own construction

3.2 Calibration results with exogenous estimates by Eurostat

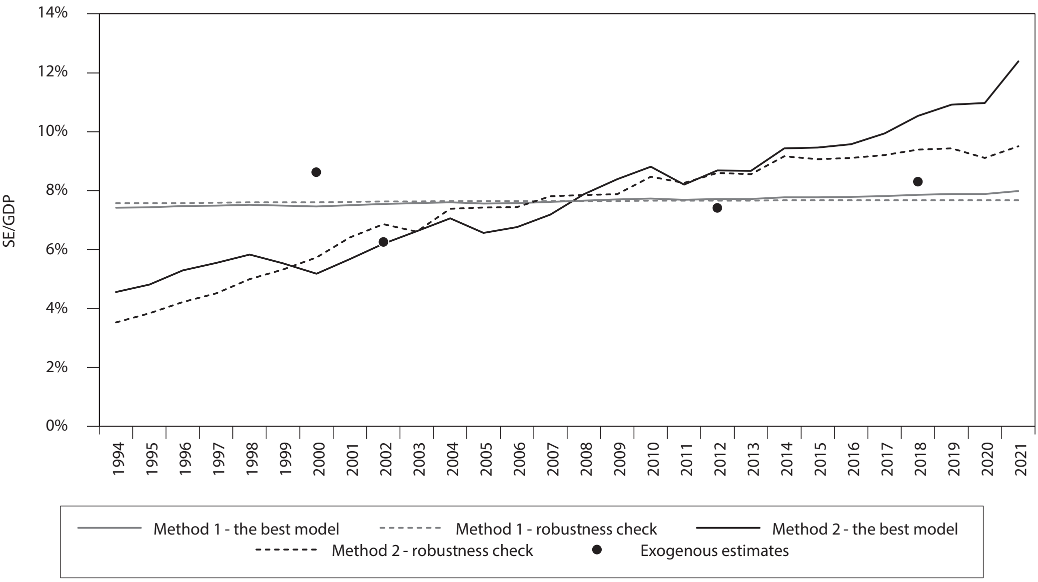
Figure 2 shows the calibration results using exogenous estimates by Eurostat. The observed types of implausible results are inverted trend, and implausibly low variability of the SE/GDP according to Method 1.

The inverted trend can be recognised in Figure 1, where both first-stage scores are declining, while in Figure 2, the resulting SE/GDPs of both methods are increasing (although very slightly for Method 1). In addition, a negative correlation between the exogenous estimates and the first-stage scores indicates the inverted trend in Method 1, and a negative correlation between the exogenous estimates and the reference indicator indicates the inverted trend in Method 2. The correlations are shown in Table 4. The implausible low variability of the SE/GDP according to Method 1 can be observed as a "too flat" time series of the resulting SE/GDP and by comparing the resulting standard deviation of the resulting SE/GDPs to the standard deviation of the first-stage scores in Table 4.

As seen from Table 4, none of the auxiliary regressions has a good fit (i.e., R^2 greater than 0.6 and coefficients statistically significant at a 10% level). The low variability of the resulting SE/GDP from Method 1 is caused mainly by the low variability of the exogenous estimates and the low correlation between the first-stage scores and the exogenous estimates.[5]

[5] Even if the correlation between the first-stage scores and the exogenous estimates were 1, the standard deviation of the resulting SE/GDP would be equal to the standard deviation of the exogenous estimates, which is 0.009.

Figure 2 Calibration results with national accounts-based exogenous estimates by Eurostat



Source: Own construction

Table 4 Decomposition of the standard deviation of the SE/GDP using exogenous estimates by Eurostat

	Method 1		Method 2	
	The best model	**Robustness check**	**The best model**	**Robustness check**
Standard deviation of the exogenous estimates	0.009	0.009	0.009	0.009
Standard deviation of the first-stage scores	0.046	0.040	0.046	0.040
Standard deviation of the first-stage scores for periods when the exogenous estimates are available	0.046	0.031	NA	NA
Correlation between the exogenous estimates and the first-stage scores	–0.165	–0.029	NA	NA
Standard deviation of the reference indicator for periods when the exogenous estimates are available	NA	NA	0.122	0.122
Correlation between the exogenous estimates and the reference indicator	NA	NA	–0.166	–0.166
Standard deviation of the SE/GDP	**0.002**	**< 0.001**	**0.021**	**0.018**
R^2 of the auxiliary regression	0.027	0.001	0.027	0.027
Significant coefficients in the auxiliary regression at a 10% level	NO	NO	NO	NO

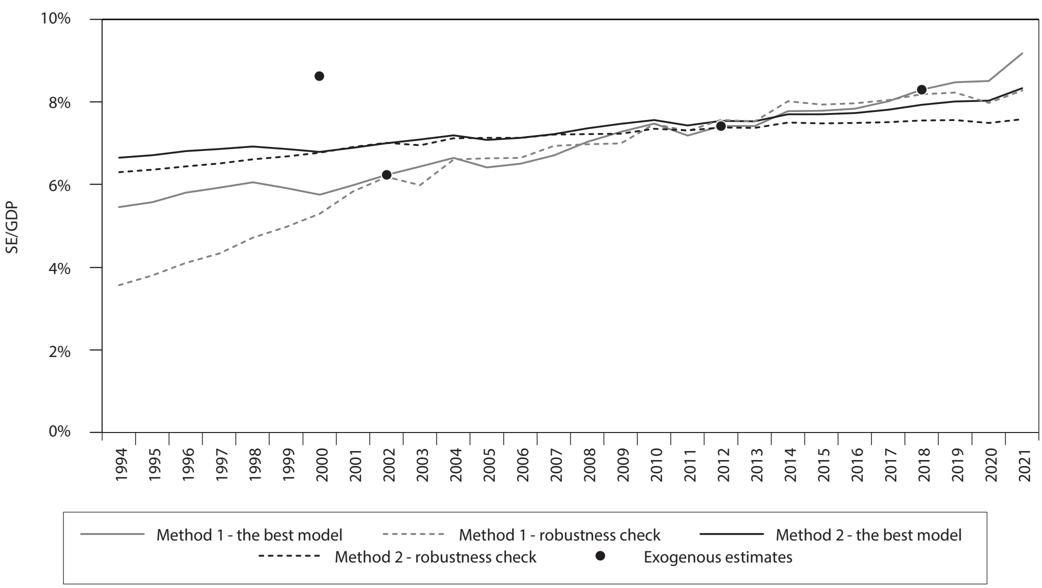
Notes: NA means that the parameter is not part of the standard deviation decomposition for the particular method.
Source: Own construction

Two things cause the much higher variability of the result of Method 2: the different methodology of the calculation and the standard deviation of the reference indicator being much higher than the standard deviation of the first-stage scores for the observations with exogenous SE estimates available. In Method 1, the low correlation between the exogenous estimates and first-stage scores deflates the resulting SE/GDP variability. In contrast, in Method 2, the low correlation between the exogenous estimates and the reference indicator inflates the resulting SE/GDP variability.

Regarding Method 1, the results for the best model and for the robustness check are similar. The results from the robustness check show that the lower variability is due to the very low correlation between the first-stage scores from the robustness check and the exogenous estimates. Regarding Method 2, slight divergences between the best model and the robustness check arise from different time paths of the first-stage scores.

Figure 2 shows that while the exogenous estimates from 2002, 2012, and 2018 have an increasing trend, the exogenous SE/GDP in 2000 is the highest and causes an overall trend of the exogenous estimates to be almost constant, which makes it an influential point in the regression. The inclusion or exclusion of the 2000 exogenous estimate notably impacts the correlation between the exogenous estimates and the first-stage scores (–0.165 with the 2000 observation and –1.000 without it for the best model), as well as the correlation between the exogenous estimates and the reference indicator (–0.166 with the 2000 observation and –0.715 without it for the best model), consequently affecting auxiliary regression coefficients and adjusted MIMIC coefficients. After excluding the 2000 observation, the standard deviation of the resulting SE/GDP from Method 1 increased for both models. In contrast, for Method 2, the SE/GDP of both models decreased.

Figure 3 Calibration results with national accounts-based exogenous estimates by Eurostat without the 2000 estimate



Source: Own construction

Figure 3 shows when the exogenous estimate from 2000 is excluded, the results of the two methods are much closer together for both the best model and the robustness check, and the SE/GDP according to Method 1 is much less flat. In addition, the R^2 of the auxiliary regressions is greater for both models and methods. The coefficients yielded by Method 1 are significant at a 10% level for both models.

Table 5 Decomposition of the standard deviation of the SE/GDP using exogenous estimates by Eurostat without the 2000 estimate

	Method 1		Method 2	
	The best model	**Robustness check**	**The best model**	**Robustness check**
Standard deviation of the exogenous estimates	0.008	0.008	0.008	0.008
Standard deviation of the first-stage scores	0.046	0.040	0.046	0.040
Standard deviation of the first-stage scores for periods when the exogenous estimates are available	0.039	0.023	NA	NA
Correlation between the exogenous estimates and the first-stage scores	–1.000	–0.990	NA	NA
Standard deviation of the reference indicator for periods when the exogenous estimates are available	NA	NA	0.120	0.120
Correlation between the exogenous estimates and the reference indicator	NA	NA	–0.715	–0.715
Standard deviation of the SE/GDP	**0.010**	**0.014**	**0.004**	**0.004**
R^2 of the auxiliary regression	0.999	0.981	0.511	0.511
Significant coefficients in the auxiliary regression at a 10% level	YES	YES	NO	NO

Notes: NA means that the parameter is not part of the standard deviation decomposition for the particular method.
Source: Own construction

Nevertheless, the inverted trend persists because the correlations are negative. Even though the exclusion of the 2000 exogenous estimate made the results of the two calibration methods much closer together, the authors do not advise excluding the influential points unless there are reasonable doubts about the reliability of the exogenous estimate in the particular period.

3.3 Calibration results with exogenous estimates from the survey

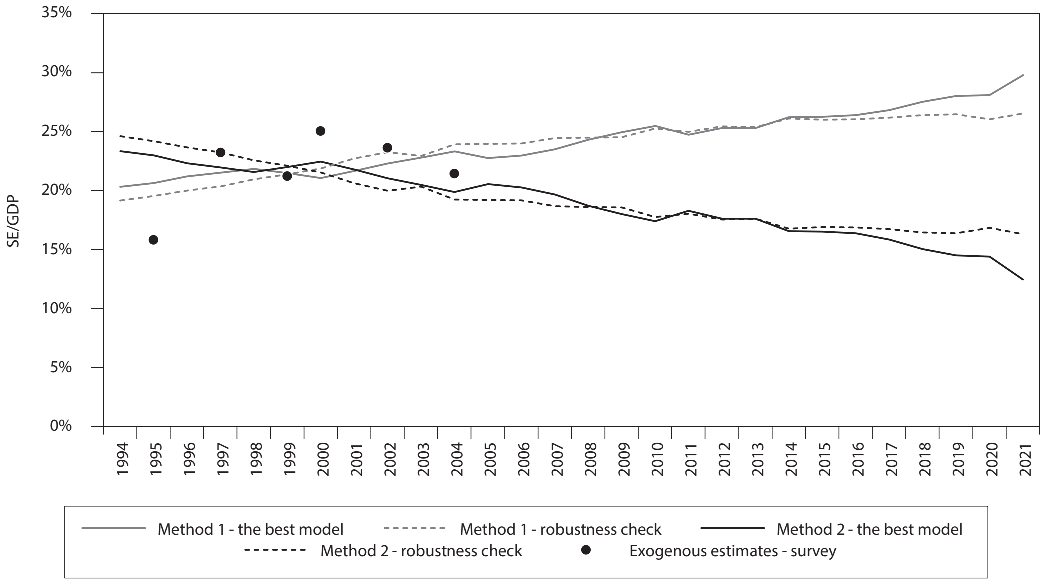
Figure 4 shows the results of calibration to survey-based estimates. The trends of the results of the two methods are opposite, which is caused by the inverted trend being present while applying Method 1 but not while applying Method 2. That is because the exogenous estimates positively correlate with the reference indicator but negatively with the first-stage scores. The exogenous estimates oscillate like a sine wave, which increases the risk of an inverted trend, especially in this case with few observations. Although the reference indicator and the first-stage scores are positively correlated (both having a global decreasing trend), the reference indicator has a strong local increasing trend between 1994 and 2002 (see Figure 1).

The auxiliary regressions have a bad fit for Method 1 for both models. The R^2 is 0.089 for the best model, and 0.267 for the robustness check, and the coefficients are not significant on a 10% level. In contrast, for Method 2, the auxiliary regression fit is much better for both models. The R^2 is 0.827, and the coefficients are statistically significant at a 10% level for both models. Generally, the results are very similar for the best model and for the robustness check.

The exogenous estimate from 1995 is an influential point that makes the trend of the exogenous estimates increasing, even though the other observations have a declining trend. The calibration results without the exogenous estimate from 1995 are shown in Figure 5.

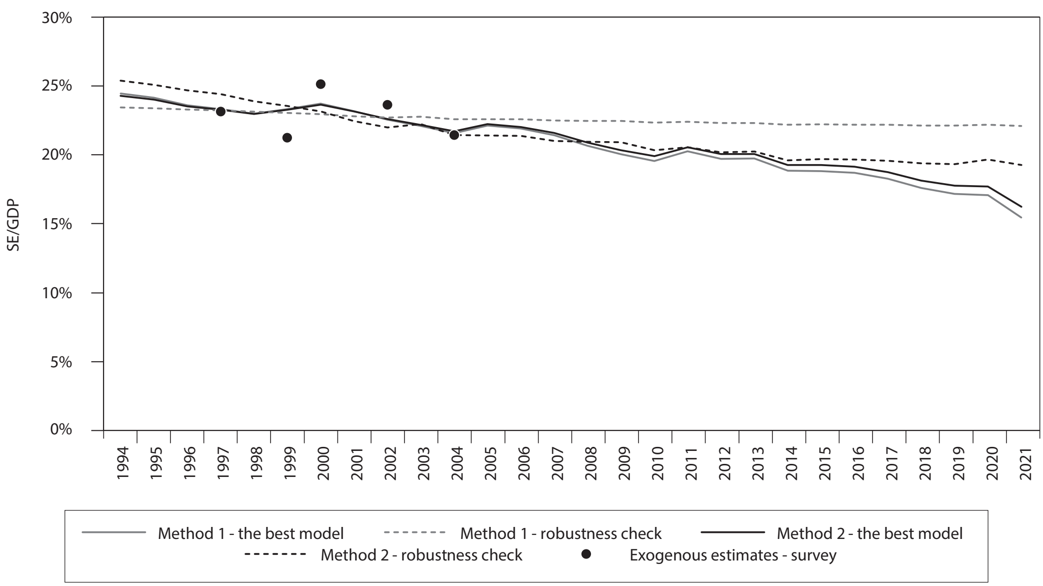
After the exclusion, the auxiliary regressions still have a bad fit for Method 1 and a slightly worse but still relatively good fit for Method 2. Nevertheless, the results of both methods are very similar and do not have an inverted trend. However, the reason for the similarity of the results is not the correlations being

Figure 4 Calibration results with survey-based exogenous estimates



Source: Own construction

Figure 5 Calibration results with survey-based exogenous estimates without the 1995 estimate



Source: Own construction

close to 1 or the standard deviations of the first-stage scores and the reference indicator being close (see Table 6). The reason is that $|corr(\hat{SE}_W^{FS\_est}, SE_W^{exog})|$ is approximately 2.5 times smaller than $\frac{1}{|corr(y_{1,W}, SE_W^{exog})|}$, and $\hat{\sigma}_{\hat{SE}_W^{FS\_est}}$ is approximately 2.5 times greater than $\sigma_{y_{1,W}}$.

However, the resulting SE/GDP decreased between 2008 and 2010 during the financial and economic crisis when it was expected to rise. The exogenous estimates do not cover this period. Therefore, the local trend between 2008 and 2010 is determined by the local trends of the exogenous estimates, the first-stage scores, and reference indicator for the observation with exogenous SE estimates available. Hypothetically, the declining trend of the estimated SE/GDP could result from a coincidence that the local trends of the exogenous estimates, the first-stage scores, and reference indicator are accordant even though the global trends differ.

The resulting SE/GDP from the robustness check using Method 1 has a significantly lower variability than the resulting SE/GDP from the best model. That is mainly because the first-stage scores used for the robustness check are much less correlated to the exogenous estimates than the first-stage scores from the best model.

Table 6 Decomposition of the standard deviation of the SE/GDP using survey-based exogenous estimates without the 1995 estimate

	Method 1		Method 2	
	The best model	**Robustness check**	**The best model**	**Robustness check**
Standard deviation of the exogenous estimates	0.014	0.014	0.014	0.014
Standard deviation of the first-stage scores	0.046	0.040	0.046	0.040
Standard deviation of the first-stage scores for periods when the exogenous estimates are available	0.015	0.023	NA	NA
Correlation between the exogenous estimates and the first-stage scores	0.529	0.162	NA	NA
Standard deviation of the reference indicator for periods when the exogenous estimates are available	NA	NA	0.037	0.037
Correlation between the exogenous estimates and the reference indicator	NA	NA	0.818	0.818
Standard deviation of the SE/GDP	**0.024**	**0.004**	**0.022**	**0.019**
R^2 of the auxiliary regression	0.280	0.026	0.670	0.670
Significant coefficients in the auxiliary regression at a 10% level	NO	NO	YES	YES

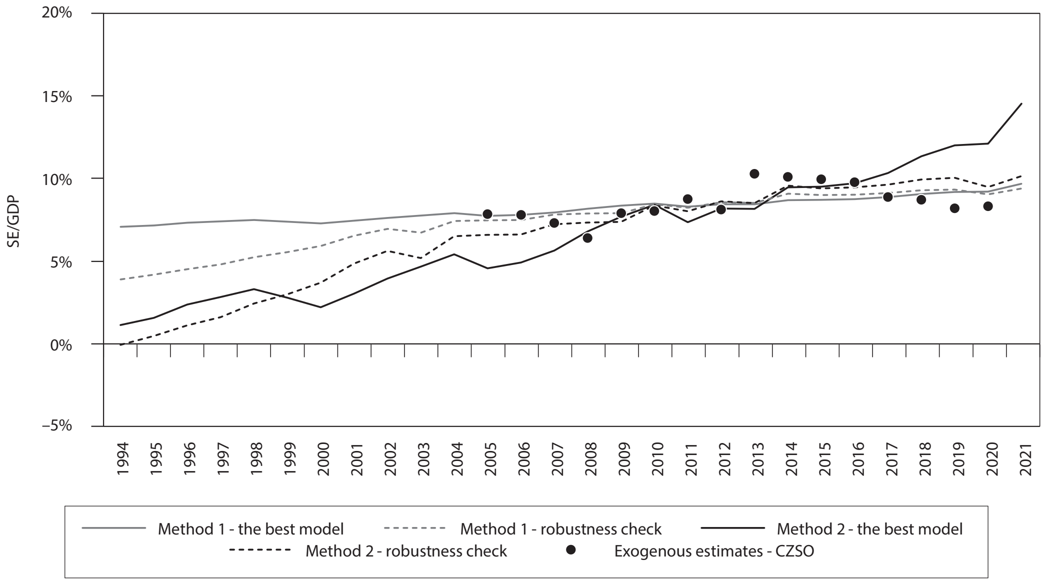
Notes: NA means that the parameter is not part of the standard deviation decomposition for the particular method.
Source: Own construction

3.4 Calibration results with exogenous estimates by CZSO

Despite having 16 years of available exogenous estimates by the CZSO, the calibration results displayed in Figure 6 are implausible. The observed types of implausibility include an inverted trend, implausibly low variability of the results for the robustness check with Method 1 for the robustness check, implausibly low values of the resulting SE/GDP from Method 2 at the beginning of the time series (and negative for the robustness check), and notable divergences in SE/GDP variabilities between Method 1 and Method 2.

The causes of the inverted trend, the implausibly low variability in Method 1, and the significant divergences in variability between resulting SE/GDP from Method 1 and Method 2 are the same as discussed in subsection 3.2. The implausibly low values from the results by Method 2 are caused by low values of

Figure 6 Calibration results with the national accounts-based estimate by CZSO



Source: Own construction

Table 7 Decomposition of the standard deviation of the SE/GDP using exogenous estimates by CZSO

	Method 1		Method 2	
	The best model	**Robustness check**	**The best model**	**Robustness check**
Standard deviation of the exogenous estimates	0.010	0.010	0.010	0.010
Standard deviation of the first-stage scores	0.046	0.040	0.046	0.040
Standard deviation of the first-stage scores for periods when the exogenous estimates are available	0.029	0.015	NA	NA
Correlation between the exogenous estimates and the first-stage scores	–0.424	–0.599	NA	NA
Standard deviation of the reference indicator for periods when the exogenous estimates are available	NA	NA	0.056	0.056
Correlation between the exogenous estimates and the reference indicator	NA	NA	–0.238	–0.238
Standard deviation of the SE/GDP	**0.007**	**0.017**	**0.036**	**0.031**
R^2 of the auxiliary regression	0.179	0.358	0.056	0.056
Significant coefficients in the auxiliary regression at a 10% level	NO	YES	NO	NO

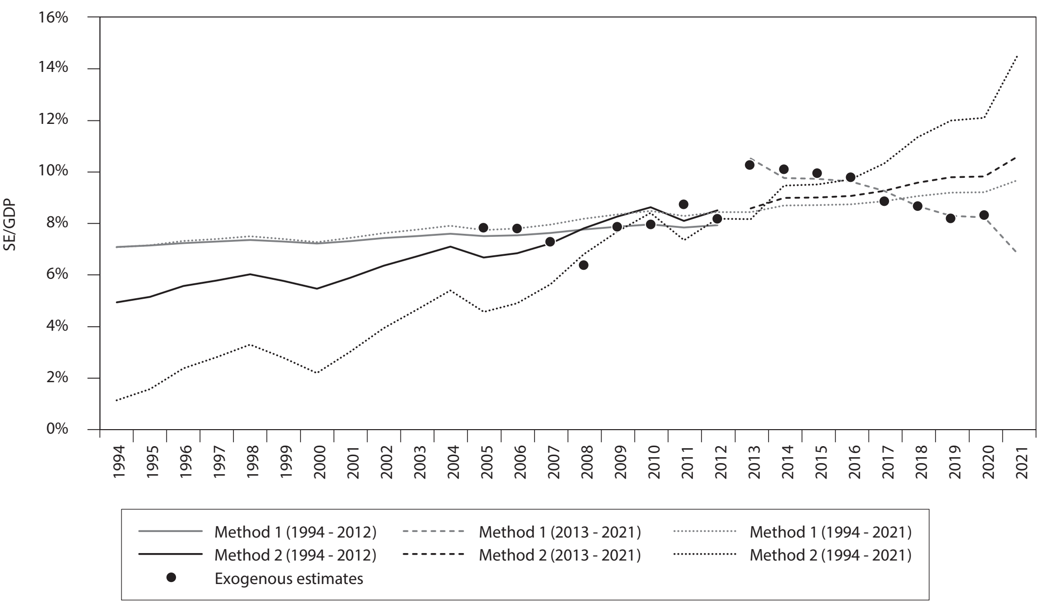
Notes: NA means that the parameter is not part of the standard deviation decomposition for the particular method.
Source: Own construction

the exogenous estimates placed at the end of the time series, and a steep upward slope of the estimated SE/GDP according to Method 2 (inverted due to the inverted trend). In addition, the low correlation between the exogenous estimates and the reference indicator inflated the resulting SE/GDP variability.

The trend of the exogenous estimates being opposite to the trends of the reference indicator and the first-stage scores for the observations with exogenous SE estimates available is likely caused by the implicit SE definition from the model not corresponding to the definition by CZSO. Moreover, Table 7 shows that the fits of auxiliary regressions for both models and methods are bad.

The results of both models show the same kinds of implausibility for both methods. Method 2 resulting SE/GDP from the robustness check is negative at the beginning of the time series. A significantly higher variability of the resulting SE/GDP from the robustness check in Method 1 is caused mainly by the lower variability of the first-stage scores for the observations with exogenous SE estimates available and the higher negative correlation between the first-stage scores and the exogenous estimates.

Figure 7 Robustness check results with exogenous estimates by CZSO



Source: Own construction

The calibration methods by Dell'Anno (2022) assume that the same economic principles apply for the entire period as for the period for which the exogenous estimates are available. The subsequent robustness check is employed to verify if this assumption holds by performing the calibrations for two periods separately. This test was performed only for the CZSO calibration since the other exogenous estimates have too few observations. The exogenous estimates were divided into two parts. The first part was 1994–2012 (the first eight observations of the exogenous estimates), and the second was 2013–2021.

Figure 7 shows that implausibly low values of the resulting SE/GDP at the beginning of the time series, according to Method 2, are no longer such an issue. However, the inverted trend is still present in all results except for Method 1 (2013–2021). Furthermore, there is still a significant divergence between the variabilities of Method 1 and Method 2 results in both periods, and an implausibly low variability for Method 1 (1994–2012). Table 8 shows that the fits of the auxiliary regressions are not good except for Method 1 (2013–2021).

As Table 8 shows, the resulting SE/GDP standard deviations are not stable in either method. The main driver is the instability of the correlations between the exogenous estimates and the first-stage scores for

Table 8 Robustness check with exogenous estimates by CZSO

	Method 1		Method 2	
	1994–2012	2013–2021	1994–2012	2013–2021
Standard deviation of the exogenous estimates	0.006	0.008	0.006	0.008
Standard deviation of the first-stage scores	0.028	0.023	0.028	0.023
Standard deviation of the first-stage scores for periods when the exogenous estimates are available	0.018	0.017	NA	NA
Correlation between the exogenous estimates and the first-stage scores	–0.258	0.958	NA	NA
Standard deviation of the reference indicator for periods when the exogenous estimates are available	NA	NA	0.065	0.044
Correlation between the exogenous estimates and the reference indicator	NA	NA	–0.247	–0.727
Standard deviation of the SE/GDP	**0.003**	**0.010**	**0.011**	**0.006**
R^2 of the auxiliary regression	0.066	0.918	0.061	0.529
Significant coefficients in the auxiliary regression at a 10% level	NO	YES	NO	YES

Notes: NA means that the parameter is not part of the standard deviation decomposition for the particular method.
Source: Own construction

Method 1 and between the exogenous estimates and the reference indicator for Method 2. The correlation between the exogenous estimates and the first-stage scores is relatively low and negative in the first period but highly positive in the second. The correlation between the exogenous estimates and the reference indicator is negative in both periods. However, it is much stronger during the second period.

The resulting SE/GDP for Method 1 (1994–2012) does not significantly differ from Method 1 (1994–2021). However, in 2013–2021, the resulting SE/GDP shows an opposite trend and is placed much higher than the resulting SE/GDP calibrated for the entire period. The SE/GDP resulting from Method 2 calibrated on separate periods has a much shallower slope than the resulting SE/GDP calibrated for the whole period. Overall, the resulting SE/GDPs calibrated for the separate periods significantly differ from the resulting SE/GDP calibrated for the whole period. Therefore, the assumption that the same economic principles apply for the entire period is likely unmet. The new methods are sensitive to the periods to which they are applied.

CONCLUSION

This paper explained and demonstrated implausible results in the SE estimation from MIMIC models. In order to convert the first-stage scores from the model to SE/GDP, they have to be calibrated using an exogenous estimate or series of exogenous estimates. However, calibration may lead to implausible results, such as an inverted trend, negative or implausibly high resulting SE/GDP, "too flat" time series of the resulting SE/GDP, or local trends not corresponding to reasonable expectations.

This paper has examined both the traditional calibration methods (i.e., methods by Giles and Tedds, 2002; Dell'Anno and Schneider, 2003; and Bajada and Schneider, 2005), and the methods newly developed by Dell'Anno (2022) and explained why these methods yielded implausible results.

The main focus was on the new methods by Dell'Anno (2022) that allow for using an extended time series of exogenous estimates (with at least two observations). Furthermore, apart from some of the traditional methods, neither the level nor the variability depends on the level of first-stage scores. However, even the new methods may lead to implausible results when specific data issues or economic phenomena are present. This paper described and demonstrated the data issues in detail, briefly mentioning the economic phenomena.

This paper concluded that one of the critical reasons for the implausible results was the misalignment of the SE definition implied by the MIMIC model and the SE definition used for the exogenous estimates. With the traditional methods, little attention is paid to the alignment, and testing is impossible if only one exogenous estimate is available. On the contrary, the new methods use auxiliary regressions whose poor fit may indicate misalignment between the SE definition implied by the model and the SE definition used for the exogenous estimates.

The alignment of the SE definitions mentioned above is essential not only for getting plausible results but also for the reliability of the resulting SE/GDP estimate. Suppose the first-stage scores based on one SE definition are calibrated to fit the exogenous estimates based on another SE definition. In that case, the resulting figure is unclear as to which definition of the SE is represented. Therefore, the reliability of the resulting SE/GDP is questionable, and it is not advisable to use it as a base for further research or policy-making implications.

When estimating the SE using the MIMIC model, the authors recommend testing the alignment of the SE definition implied by the model with the SE definition used for the exogenous estimates using either comparison of trends, correlation, or regression. If a misalignment is discovered, either another series of exogenous estimates should be used, or the MIMIC model should be redeveloped so that the latent variable reflects the exogenous estimates more clearly.

The authors do not believe that a more advanced calibration method can solve the issue of implausible results. No matter how good the method is, making reliable SE estimates from unreliable or incompatible data is impossible. Nevertheless, more advanced calibration methods may be able to make the resulting SE/GDP fit the exogenous estimates better or may not have such strict assumptions as the methods discussed in this paper.

Further research on this topic could concentrate on applying the new calibration methods to more countries to validate the methods' applicability and possibly discover more types of implausible results. Furthermore, alternative data sources or more advanced statistical techniques could be applied to improve the accuracy of SE estimates. Another suggestion for future research is to explore in more detail economic phenomena that may lead to implausible results and link implausible results to economic theories.

ACKNOWLEDGEMENTS

For helpful comments and suggestions, we thank Friedrich Schneider,[6] Roberto Dell'Anno,[7] and two anonymous reviewers.

The article was prepared as one of the outputs of a research project of the Faculty of Finance and Accounting at the Prague University of Economics and Business „Dynamics of real and financial variables in the context of monetary and macro/microprudential policy" registered by the Internal Grant Agency of Prague University of Economics and Business under the registration number F1/14/2023.

References

BAJADA, C., SCHNEIDER, F. (2005). The Shadow Economies of the Asia-Pacific. *Pacific Economic Review*, 10(3): 379–401.
BOLLEN, K. A. (1989). *Structural Equations with Latent Variables.* Wiley. ISBN 978-1-118-61917-9
BREUSCH, T. (2005). *Estimating the Underground Economy using MIMIC Models.* EconWPA.

[6] Johannes Kepler University of Linz, Austria.
[7] University of Salerno, Italy.

BROWNE, M. W., CUDECK, R. (1993). *Testing Structural Equation Models.* SAGE. ISBN 978-0-8039-4507-4
BUEHN, A., SCHNEIDER, F. (2008). *MIMIC Models, Cointegration and Error Correction: An Application to the French Shadow Economy.* SSRN Scholarly Paper, Rochester, NY: Social Science Research Network.
CZECH STATISTICAL OFFICE. (2023). *Non-observed economy in the Czech Republic.* Prague: CZSO.
DELL'ANNO, R. (2007). The Shadow Economy in Portugal: An Analysis with the Mimic Approach online [online]. *Journal of Applied Economics*, 10(2): 253–277. <https://doi.org/10.1080/15140326.2007.12040490>.
DELL'ANNO, R. (2022). Measuring the unobservable: estimating informal economy by a structural equation modeling approach [online]. *International Tax and Public Finance.* <https://doi.org/10.1007/s10797-022-09742-0>.
DELL'ANNO, R., SCHNEIDER, F. (2003). The Shadow Economy of Italy and other OECD Countries: What Do We Know? [online]. *Journal of Public Finance and Public Choice*, 21(2–3): 97–120. <https://doi.org/10.1332/251569203X15668905422009>.
EUROSTAT. (2005). *Eurostat's Tabular Approach to Exhaustiveness* [online]. 5th Meeting of the GNI Committee. <https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=739814884&fo=0&ext=intconsult>.
EUROSTAT. (2023). *Compilation of GNI – Czech Republic 2018.* CIRCABC database.
FEIGE, E. L. (2016). Professor Schneider's Shadow Economy (SSE): What Do We Really Know? A Rejoinder [online]. *Journal of Tax Administration*, 2(2). <http://jota.website/index.php/JoTA/article/view/109/78>.
GILES, D. E. A., TEDDS, L. M. (2002). *Taxes and the Canadian Underground Economy. Canadian Tax Foundation* [online]. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1002475>.
HANOUSEK, J., PALDA, F. (2006). Vývoj daňových úniků za pomoci markovských řetězců. *Czech Journal of Economics and Finance*, 56(3–4): 127–151.
HASSAN, M., SCHNEIDER, F. (2016). *Modelling the Egyptian Shadow Economy: A Currency Demand and A MIMIC Model Approach* [online]. <https://doi.org/10.13140/RG.2.1.2296.6804>.
KIRCHGÄSSNER, G. (2016). On Estimating the Size of the Shadow Economy [online]. *German Economic Review*, 18(1): 99–111. <https://doi.org/10.1111/geer.12094>.
MAENHOUT, M. A. (2016). *The Underground Economy: Evaluation and Replication of Schneider's Estimates.* Dissertation Summary, Brussels: University of Leuven.
MARMORA, P., MASON, B. J. (2021). Does the shadow economy mitigate the effect of cashless payment technology on currency demand? Dynamic panel evidence [online]. *Applied Economics*, 53(6): 703–718. <https://doi.org/10.1080/00036846.2020.1813246>.
SCHNEIDER, F., BUEHN, A., MONTENEGRO, C. E. (2010). *Shadow Economies All Over The World : New Estimates For 162 Countries From 1999 To 2007.* The World Bank, Policy Research Working Papers.
NCHOR, D. (2021). Shadow economies and tax evasion: The case of the Czech Republic, Poland and Hungary [online]. *Society and Economy*, 43(1): 21–37. <https://doi.org/10.1556/204.2020.00029>.
OYIBO, P. V., SCHNEIDER, F. (2022). How large is the size of Côte d'Ivoire's informal sector? A MIMIC approach. *Theoretical and Applied Economics*, Vol. XXIX, No. 4(633): 205–216.

ANNEX A – COMPLETE LIST OF CAUSES

Table A1 Complete list of causes for model building in Section 2

Cause	Group
NPL ratio of business loans	Economic situation of businesses
Real total Gross fixed capital formation	Economic situation of businesses
Real average gross wage in CZK	Economic situation of households
Real amount of loans to households	Economic situation of households
Real amount of loans to households per capita	Economic situation of households
Poverty headcount ratio at $6.85 a day (2017 PPP) (% of population)	Economic situation of households
Real amount of consumer loans to households	Economic situation of households
Real amount of consumer loans to households per capita	Economic situation of households
NPL ratio of loans to households	Economic situation of households
NPL ratio of consumer loans	Economic situation of households

Table A1 (continuation)

Cause	Group
Real amount of Non-bank consumer loans	Economic situation of households
Real amount of Non-bank consumer loans per capita	Economic situation of households
Number of housing allowance recipients	Economic situation of households
Real amount of housing allowances paid	Economic situation of households
Real amount of housing allowances paid per recipient	Economic situation of households
Percentage of people who receive housing allowances	Economic situation of households
Amount of loans to households per capita/average wage	Economic situation of households
Amount of consumer loans per capita/average wage	Economic situation of households
Amount of non-bank consumer loans per capita/average wage	Economic situation of households
Percentage of people with only primary education or no education (%)	Education
Percentage of people with high school education without maturita (%)	Education
Percentage of people who have high school education with maturita (%)	Education
Percentage of people who have university education (%)	Education
Personal income tax revenue/GDP (%)	Education
Percentage of people who have maturita exam	Education
Government consumption/GDP (%)	Freedom
Fraser Institute Freedom index	Freedom
Corruption index	Freedom
Economic freedom (part of the Heritage Index)	Freedom
Business freedom (part of the Heritage Index)	Freedom
Fiscal freedom (part of the Heritage Index)	Freedom
Regulation Quality Index from the World Bank	Freedom
Government Effectiveness Index from the World Bank	Freedom
Unemployment rate (%)	Job market
Self employment rate (%)	Job market
Inflation measured by CPI (%)	Others
Public social spending/GDP (%)	Others
Net migration	Others
Number of immigrants	Others
Indirect tax revenue/GDP (%)	Taxation
Tax revenue + social security revenue/GDP (%)	Taxation
Tax revenue without social security/GDP (%)	Taxation
Tax revenue from personal income and indirect taxes/GDP (%)	Taxation
Revenue from indirect taxes/total tax revenue	Taxation
Revenue from the personal income tax/total tax revenue	Taxation

Source: Own construction

ANNEX B – DECOMPOSING THE STANDARD DEVIATIONS OF THE RESULTING SE/GDP FROM THE NEW METHODS

For Method 1, the formula for calculating the resulting SE/GDP in Formula (15) can be rewritten as in Formula (23) which follows from Formulas (12) and (14):

$$ {}^{est\_1}\widehat{SE}_t = \hat{\rho}_0 + \hat{\rho}_1 \widehat{SE}_t^{FS\_est}, \tag{23} $$

where ${}^{est\_1}\widehat{SE}_t$ is the SE/GDP estimated using Method 1, $\hat{\rho}_0$ and $\hat{\rho}_1$ are the OLS estimates of coefficients from Formula (13), and $\widehat{SE}_t^{FS\_est}$ are the first-stage scores. The standard deviation of the resulting SE/GDP is:

$$ \sigma_{est\_1\widehat{SE}} = |\hat{\rho}_1| \sigma_{\widehat{SE}^{FS\_est}}, \tag{24} $$

where $\sigma_{\widehat{SE}^{FS\_est}}$ is the standard deviation of the first-stage scores. $\hat{\rho}_1$ is a coefficient from a simple linear regression where the exogenous estimates SE_W^{exog} are the explained variable and $\widehat{SE}_W^{FS\_est}$ is the explanatory variable. Therefore, it can be calculated as:

$$ \hat{\rho}_1 = corr(\widehat{SE}_W^{FS\_est}, SE_W^{exog}) \frac{\sigma_{SE_W^{exog}}}{\sigma_{\widehat{SE}_W^{FS\_est}}}, \tag{25} $$

where $corr(\widehat{SE}_W^{FS\_est}, SE_W^{exog})$ is correlation between the first-stage scores and the exogenous estimates, $\sigma_{SE_W^{exog}}$ is the standard deviation of the exogenous estimates, and $\sigma_{\widehat{SE}_W^{FS\_est}}$ is the standard deviation of the first-stage scores for observations with available exogenous estimates. Therefore,

$$ \sigma_{est\_1\widehat{SE}_t} = \left| corr(\widehat{SE}_W^{FS\_est}, SE_W^{exog}) \frac{\sigma_{SE_W^{exog}}}{\sigma_{\widehat{SE}_W^{FS\_est}}} \right| \sigma_{\widehat{SE}^{FS\_est}} = |\sigma_{SE_W^{exog}}| \frac{\sigma_{\widehat{SE}^{FS\_est}}}{|\sigma_{\widehat{SE}_W^{FS\_est}}|} |corr(\widehat{SE}_W^{FS\_est}, SE_W^{exog})|, \tag{26} $$

where $\sigma_{est\_1\widehat{SE}_t}$ is the standard deviation of the resulting SE/GDP from Method 1. Since standard deviations cannot be negative, the $\sigma_{SE_W^{exog}}$ and $\sigma_{\widehat{SE}_W^{FS\_est}}$ can be written without absolute values.

For Method 2, the formula for calculating the resulting SE/GDP in Formula (19) can be rewritten as in Formula (27) which follows from Formulas (12) and (17):

$$ {}^{est\_2}\widehat{SE}_t = {}^{\Delta\mu\_est}\hat{\gamma}_0^* + \frac{\widehat{SE}_t^{FS\_est}}{{}^{REG}\hat{\lambda}_1}, \tag{27} $$

where ${}^{\Delta\mu\_est}\hat{\gamma}_0^*$ is the intercept calculated according to Formula (18), and ${}^{REG}\hat{\lambda}_1$ is the OLS estimate of the parameter ${}^{REG}\lambda_1$ from Formula (16). Therefore, its standard deviation is:

$$ \sigma_{est\_2\widehat{SE}_t} = \frac{\sigma_{\widehat{SE}^{FS\_est}}}{|{}^{REG}\hat{\lambda}_1|}, \tag{28} $$

where $\sigma_{est\_2\hat{SE}_t}$ is the standard deviation of the resulting SE/GDP from Method 2. ${}^{REG}\hat{\lambda}_1$ an OLS estimate of a coefficient from the regression described in Formula (16). Using the same rules as for $\hat{\rho}_1$ in Formula (25), ${}^{REG}\hat{\lambda}_1$ can be rewritten as:

$$ {}^{REG}\hat{\lambda}_1 = corr(Y_{1,W}, SE_W^{exog}) \frac{\sigma_{Y_{1,W}}}{\sigma_{SE_W^{exog}}}, \tag{29} $$

where $corr(Y_{1,W}, SE_W^{exog})$ is the correlation between the reference indicator and the exogenous estimates, $\sigma_{Y_{1,W}}$ is the standard deviation of the reference indicator for observations with available exogenous estimates. The standard deviation of the resulting SE/GDP according to Method 2 is:

$$ \sigma_{est\_2\hat{SE}_t} = \frac{\sigma_{\hat{SE}^{FS\_est}}}{\left| corr(Y_{1,W}, SE_W^{exog}) \frac{\sigma_{Y_{1,W}}}{\sigma_{SE_W^{exog}}} \right|} = \left|\sigma_{SE_W^{exog}}\right| \frac{\sigma_{\hat{SE}^{FS\_est}}}{\left|\sigma_{Y_{1,W}}\right|} \frac{1}{\left|corr(Y_{1,W}, SE_W^{exog})\right|}. \tag{30} $$

Since standard deviations cannot be negative, the $\sigma_{SE_W^{exog}}$ and $\sigma_{Y_{1,W}}$ can be written without absolute values.

ANNEX C – DERIVING THE EQUALITY CONDITION OF RESULTS OF CALIBRATION METHOD 1 AND METHOD 2

For Method 1, the derivation is based on Formula (23). Formula (25) is used to derive $\hat{\rho}_1$. $\hat{\rho}_0$ is an OLS estimate of a constant from a simple linear regression where the exogenous estimates SE_W^{exog} are the explained variable and $\hat{SE}_W^{FS\_est}$ is the explanatory variable. Therefore, it can be calculated as:

$$ \hat{\rho}_0 = Mean(SE_W^{exog}) - \hat{\rho}_1 Mean(\hat{SE}_W^{FS_{est}}) = Mean(SE_W^{exog}) - corr(\hat{SE}_W^{FS_{est}}, SE_W^{exog}) \frac{\sigma_{SE_W^{exog}}}{\sigma_{\hat{SE}_W^{FS\_est}}} Mean(\hat{SE}_W^{FS_{est}}), \tag{31} $$

where $Mean(SE_W^{exog})$ is the mean of the exogenous estimates, and $Mean(\hat{SE}_W^{FS_{est}})$ is the mean of the first-stage scores for the observations with exogenous SE estimates available.

Therefore, the SE/GDP according to Method 1 can be rewritten as:

$$ \begin{aligned} {}^{est\_1}\hat{SE}_t = {} & Mean(SE_W^{exog}) - corr(\hat{SE}_W^{FS_{est}}, SE_W^{exog}) \frac{\sigma_{SE_W^{exog}}}{\sigma_{\hat{SE}_W^{FS_{est}}}} Mean(\hat{SE}_W^{FS_{est}}) + \\ & + corr(\hat{SE}_W^{FS_{est}}, SE_W^{exog}) \frac{\sigma_{SE_W^{exog}}}{\sigma_{\hat{SE}_W^{FS_{est}}}} \hat{SE}_t^{FS_{est}} = \\ & Mean(SE_W^{exog}) + corr(\hat{SE}_W^{FS_{est}}, SE_W^{exog}) \frac{\sigma_{SE_W^{exog}}}{\sigma_{\hat{SE}_W^{FS_{est}}}} (\hat{SE}_W^{FS_{est}} - Mean(\hat{SE}_W^{FS_{est}})). \end{aligned} \tag{32} $$

For Method 2, the derivation is based on Formula (27). ${}^{REG}\hat{\lambda}_1$ is derived in Formula (29), and ${}^{\Delta\mu\_est}\hat{\gamma}_0^*$ is calculated in Formula (18). Therefore, the resulting SE/GDP according to Method 2 can be calculated as:

$$ {}^{est\_2}_{\square}\hat{SE}_t = Mean(SE_W^{exog}) - Mean\left(\frac{\hat{SE}^{FS_{est}}_{\square}}{{}^{REG}_{\square}\hat{\lambda}_1}\right) + \frac{\hat{SE}^{FS_{est}}_W}{corr(Y_{1,W}, SE_W^{exog})\,\frac{\sigma_{Y_{1,W}}}{\sigma_{SE_W^{exog}}}} \,. \tag{33} $$

After simplifications of the equations, the equality ${}^{est\_1}_{\square}\hat{SE}_t = {}^{est\_2}_{\square}\hat{SE}_t$ follows when:

$$ \frac{1}{\sigma_{\hat{SE}_W^{FS\_est}}}\left|corr(\hat{SE}_W^{FS\_est}, SE_W^{exog})\right| = \frac{1}{\sigma_{Y_{1,W}}}\,\frac{1}{\left|corr(Y_{1,W}, SE_W^{exog})\right|} \,. \tag{34} $$