

## Referee report

I have both major and minor concerns about the paper.

Major conceptual issues to consider

Many things are used interchangeably and are sometimes confusing. Most importantly, each of the following 5 notions need a precise definition

- tax effort
- tax capacity
- tax collecting efficiency
- tax potential
- tax behavior

Is it worth considering/contrasting efficient tax collection and tax capacity? Are they the same? It is stated that tax capacity depends on tax rate, tax management and inspection. It would be helpful to precisely define what is the focus of this paper.

There is no theory or theoretical considerations about the tax, tax effort, potential etc. For example, the tax rate is a part of the equation. If fiscal pressure is considered (as is mentioned several times by authors), the greater is the fiscal pressure, the larger should be the tax rate, but that would in turn increase the social pressure, which would reduce the potential of tax increases.

Following up on lack of theoretical considerations, why is there a tax frontier? What theory is stating that there is a kind of a "tax production" function?

When considering taxes at this level of aggregation, it is worth differentiating between corporate and private taxes. Sometimes low corporate taxes are used as an attraction for businesses, which create jobs.

Major empirical issues to consider

What are the determinants of the frontier and inefficiency? Are they commonly used in the literature?

Thus, some structure is required. What regions do is

- manage
- monitor
- inspect

taxes. Which of these three is(are) causing inefficiency. Which of the factors that you consider on page 16 refer to which area?

The same goes for the frontier. You provide a regression analysis to choose the (groups of) variables, but relating to existing literature will be useful.

The conclusions are limited as the TAX variable lumps different type of taxes. Is there a possibility to consider multiple outputs?

Why do you use feasible GLS, order-alpha and order-m f frontiers approaches? The former does not account for inefficiency, while the latter does not consider heterogeneity and panel structure. They are not comparable. If xtsc is correct, sfa is wrong as the distribution of errors does account for what xtsc accounts.

n=15 is very small for SFA, which is a consistent estimator.

Minor empirical issues

zones in figure 1 are arbitrary.

Page 3: first paragraph is confusing, is it an absolute value, it is a ratio. More precision is desirable.

How is your work related to that of Medina (2012)?

Would it be useful to contrast two systems: one that is considered and the one by Basque Country and Navarre?

Regression in appendix is in levels but in table 2 is in logs. It is useful to clarify this and preferable to use a unified approach.

Page 11: what is "homoskedastic sample"?

How is QMANAG related to inefficiency?

Page 15: at the bottom it is stated the coefficients are similar, but they are not.

What does exactly column XTSCC in table 3 mean? How was it calculated?