

Let it snow! Let it snow! Let it snow!
Estimating cocaine production using a novel
dataset based on reported seizures of
laboratories*

Riccardo Leoncini[†]
Francesco Rentocchini[‡]

Abstract

Background: data on cocaine market appear to be inconsistent, as they tend to show declining prices vis a vis a steady or increasing demand and a declining supply. This paper proposes an explanation by providing evidence of an under-estimation of the supply of cocaine.

Methods: we put forward a conservative estimate of cocaine production in Colombia for year 2008 by means of reported seizures of laboratories made by the counteracting organisations operating in the Colombian territory.

*A preliminary version of this paper was presented to the Conference “Cento anni di controllo internazionale sulle droghe”, Forlì, 21-22 April 2009, to the ESSD Annual Conference, Belfast, 24-26 September 2009, and to the ISSDP Annual Conference, Santa Monica (CA), 15-16 March 2010. The authors are grateful to the participants for comments and suggestions. Sandro Donati was instrumental to the decision to commit ourselves to the difficult task of elaborating these new and, hopefully, provocative ideas: we thank his intelligence, patience and enthusiasm. We would like to thank three anonymous referees and the editor of the journal for very useful and detailed comments. Francesco Rentocchini acknowledges funding from the Autonomous Province of Trento (Grant 2009 Outgoing: OVERNET). The usual caveats apply.

[†]Department of Economics, University of Bologna (I) and CERIS-CNR, Milan (I) riccardo.leoncini@unibo.it

[‡]Department of Economics, University of Trento (I) and INGENIO-UPV, CSIC, Valencia (E) francesco.rentocchini@economia.unitn.it

Results: our estimate is at least twice than that declared in the official statistics, although we are careful to keep all of our variables to their minimum boundary values. We conclude that our methodology could prove to be a very useful tool especially if used in parallel with the standard ones. Moreover, its characteristics (cheapness, ease of use and possibility to adopt it all over the world) make of it a powerful instrument for the counteracting activity against cocaine production.

Introduction

It is now a commonly agreed fact that scholars do not agree on the reliability of the data utilised in the studies on drug-related issues.

The illegal nature of the market for drugs implies, first of all, that direct measurement of the phenomenon is very difficult, and that, by relying on indirect measurements, at each step some assumptions on latent variables have to be made, and each assumption, far from being neutral, comes at a price (Thoumi, 2005). For instance, there is not a proper market indicating quality or scarcity as function of price “only”. Moreover, total quantity of hectares cultivated has to be estimated and this estimate makes use of a set of correction factors and transformation coefficients that increases the uncertainty surrounding the final estimate (see, for instance, the very high number of caveats accompanying Table 19 of UNODC (2010, p. 162) on “Global illicit cultivation of coca bush and production of coca leaf and cocaine, 1995-2009”). This adds a bias for each step, which results in huge distortions in the final data produced. Moreover, being the methodology not well established and therefore ‘constant’, even the same data can (and frequently do) lead to very different results and interpretations. To this adds the fact that not all the final users of so-produced data “somehow seek incontestable figures that may be used and exploited, dismissing uncertainties and statistical weaknesses, as well as the figures “conceptual complexities” (Thoumi, 2005, p. 186).

From the methodological point of view, it can be shown that by taking into account the true dynamic nature of such a market, the standard results of comparative static economic analysis (Caulkins et al., 2006) can be overturned. In particular, Nell (1994) shows that, by taking into account a truly dynamical approach, irreversibilities arise such that the development process of the counteracting activities will likely enter phases of lock-in. The net

result is thus not only that attempts at eradicating cultivations will fail, but, even worse, these activities are likely to expand the market. Furthermore, the effects of socio-political activities (Reuter, 2001; Thoumi, 2002) that overlap with the economic side, have surely an important role, that is almost always underplayed by the various models.

As a result of these difficulties, data on supply and demand are thoroughly inconsistent with price dynamics, and thus frequently ad-hoc explanations are put forward. In order to cope with this issue, we try to deal with cocaine production by advocating to a completely different methodology, in order to have an alternative estimate of it consistent with the global picture. We aim thus to furnishing some additional tools to complement the more standard ones. With this purpose in mind, we have collected on a daily basis all the data on seizures of laboratories producing cocaine performed in 2008 by all the counteracting organisations operating on the field in Colombia (i.e. all the legal organisations deliberately intervening against drug trafficking in order to reduce its force or neutralize it). The setting for our study is Colombia, a good laboratory since almost all of the cocaine consumed in North America and Europe comes from the Andean nations of Bolivia, Colombia and Peru: according to the Colombia coca survey for 2008 (UNODC, 2009a,b), Colombia accounts for more than the 50% of world cocaine production (its share was more than 60% in 2007). In so doing, we aim at showing that it is possible to obtain an estimate of cocaine production in Colombia based on a dataset never used before, and that can help in gaining more insight in understanding such a complex phenomenon.

Demand and supply in the cocaine market and limits to standard evaluations methods

In analysing official statistics on cocaine market, some awkward stylised facts emerge relative to the price/production relationship. Indeed, we observe a long-term decreasing global trend in the price of cocaine (bottom panel of Figure 1), vi-a-vis a more or less constant level of cocaine purity (top panel of Figure 1).

Although there have been some fluctuations over the years, the downward trend is quite evident, as to every spike in price, consumers adjusted to changed availability and suppliers reacted to higher profits presumably with

new entries (Figure 1 is taken from Fries et al. (2008), who produced the report “under contract to the ONCDP” and base their estimates on STRIDE data).

A steep decline in price would be compatible with either a decline in demand or an increase in supply or both. As far as demand is concerned, official data are quite clear in contrasting a declining demand in the US to an increasing demand in the rest of the world, thus consolidating an overall increase in world demand for cocaine (UNODC, 2010). Also the data on prevalence show a stable patterns in USA since 2000 (with a slight decline in the last two years), and an increasing pattern for some EU countries and a stable one for the rest. However, all these countries show very consistent increase in prevalence during the last 10 years. The global values of prevalence (UNODC, 2010, p. 68) confirms a long-term increase since 2000, with an overall stability in the last 3 years. Hence, data on single regions and/or nations reveal unequivocally an increase in all the indicators regarding world prevalence and demand.

As far as supply side is concerned, all the available official data point to a decrease in supply as result of a set of exogenous shocks in production (i.e. increasing eradication and aerial spraying of cocaine plants, together with increasing seizures of laboratories) and distribution (i.e. increasing activities against cross-border trafficking and increasing pressure on retail markets).

As result of this counteracting activities, official data (i.e. data from UNODC (United Nations Office on Drugs and Crime)) record a large decrease in coca cultivation in Colombia (58% less in the period 2000-2009), which gives a total production of pure cocaine of 450 metric tons in 2008 (and the decrease is expected also in 2009), which is the lowest level since 2002. This is mainly due to a strong decrease in the area under coca cultivation (which decreased by 16% with respect to 2008 to a total of 68,000 ha). The programs of manual eradication and aerial spraying show a decrease in 2009 after consistent increases in the three previous years. Also seizures of laboratories have recorded a steady growth to reach the peak since 1990 of more than 500 metric tons in 2008 (UNODC, 2010).

In the light of these data, the decline in the price of cocaine is quite puzzling, to the point that, on the one side, even UNODC is forced to concede that something must have been happening to justify this conundrum, such as, “the amount used by each consumer must have declined significantly”, or “laboratory efficiency may have increased stronger than is currently reflected in cocaine production estimates” (UNODC, 2010, p. 68), but both

statements are hard to prove as there are no official data available. On the other side, scholars are forced to investigate the issue by means of ad-hoc explanations. For instance, once some of the most violent dealers (i.e. the dealers who are more likely to be arrested) are removed from the market, a decrease in price can be observed (rather than an increase) because the exit of a dealer does not reduce supply (because of the existence a cue of new entrants) and, moreover, the likely entrance of less violent dealers will reduce external costs for other dealers, in turn causing a reduction in price (Caulkins et al., 2006). Similarly, another model advocates to globalisation and to differentiated types of dealers to show that globalisation reduces prices through the joint action of decreasing intermediation margins and an inflow of newcomers low-level dealers attracted by high profits gained by upper-level dealers (Costa-Storti and De Grauwe, 2009; Costa Storti and De Grauwe, 2009).

In order to make sense of these puzzling phenomenon, we propose an alternative explanation. Our claim is that official statistics from UNODC and ONDCP (US Government White House Office of Drug Control Policy), despite their huge efforts and the important results obtained in counteracting cocaine production, are likely to under-estimate the real production of cocaine in Colombia. In particular, the procedure used for the interpretation and processing of satellite images of coca crops to produce an estimate of potential cocaine production in Andean regions have some inevitable drawbacks, as suggested by the extant literature (Thoumi, 2005; Mejia and Posada, 2008). First, satellite images used by official agencies providing official statistics on cocaine production in Colombia do not cover the whole country. Although this can be explained by the fact that some areas do not have a soil susceptible of cocaine plant cultivation, this can actually miss some important areas where coca is actually grown. For example, the departments of Meta and Vichada are only partially covered by satellite images available to UNODC for the estimate of potential production of cocaine in Colombia in 2008 (UNODC, 2009a, p. 88). However, information on seizures of laboratories contained in our study reveal that a relevant number of highly productive laboratories are actually seized in these two areas (see Table 3). Second, an estimation of the potential production of pure cocaine inevitably requires several corrections (e.g. cloud cover, gaps, spraying, etc.) and assumptions (e.g. yield per hectare, transformation coefficients from cocaine base to cocaine hydrochloride, etc.) that can introduce a certain level of uncertainty on the final estimate of pure cocaine produced (UNODC, 2009a).

Finally, increasing evidence show how coca growers are likely to induce a strong bias in the figures of coca cultivation gathered through satellite images by intermingling coca crops with legal crops to avoid both monitoring and eradication (Thoumi, 2005; Mejia and Posada, 2008).

Although a very intense activity of verification through field work is carried out by the official agencies, still a number of drawbacks are sources of bias in the statistics on cocaine production. Therefore, our aim is to put forward a completely new dataset, with which we aim at complementing existing estimates on the amount of cocaine produced in Colombia in 2008. Our dataset allows us to supply a data on production which fits with the available picture of the cocaine market, thus supplying an explanation to the decrease of both retail prices and intermediation margins, in this way complementing the hypothesis developed by Caulkins et al. (2006) and Costa-Storti and De Grauwe (2009); Costa Storti and De Grauwe (2009). Moreover, our methodology can be used to complement the available statistics in order to get more reliable data to comprehend this phenomenon.

The dataset

The data we use in this paper have been collected by Sandro Donati, on a daily basis, from the press releases published on the websites, of all the main organisations — National Police (Policía Nacional), National Army (Ejército Nacional), Colombian Navy (Armada Nacional), Security Service Agency (Departamento Administrativo de Seguridad, DAS) — through all their main articulations (Divisions, Brigades, Battalions) covering more than 600 different territorial units performing counteracting activities on the Colombian territory. In this way, we have been able to gather data detailing each single seizure of laboratories for the production of cocaine hydrochloride. Thus, we have information on the final step of the manufacturing process of cocaine which yields the final product with an average purity level close to 100%. For each seizure a press release is published by the organisation that accomplished the operation reporting several information: the geographical location where the seizure took place; the number of people involved; the quantity of either cocaine leaves, cocaine base or cocaine hydrochloride; the quantity of both solid (e.g. potassium permanganate) and liquid (e.g. kerosene) chemical precursors; the number and the type of instruments seized (such as microwave, scales, compressors, etc.). For a certain number of laboratories seized, an

expert evaluation of the monthly production potential of that particular laboratory is supplied (normally by a chemist following the counteracting unit).

Several procedures have been put forward to clean the dataset. First of all, the data have been cross-checked in order to avoid duplication due to different counteracting units reporting the same seizure. Furthermore, we accounted for the high level of clusterisation of laboratories seized that we noticed from time to time. This was mainly due to particularly complex operations set up by a particular counteracting unit which resulted in expeditions lasting for several days in areas far from the home base. In all of these cases, the counteracting unit was reporting the information only once the mission was accomplished thus providing severe outliers in our dataset. To overcome such a problem, we looked for collateral information (e.g. press releases from local and national newspapers, details in the press releases from the counteracting organisations, etc.) which allowed us to split the data across different laboratories. Finally, in order to avoid replications that might have gone unnoticed, in case of seizures of laboratories occurring during the same day in the same region, we opted for including only one.

After the cleaning-up, the dataset comprises 306 counteracting operations carried out during the period between the 1st of January and the 31st of December 2008. In this time span, the 306 operations led to the seizures of 333 laboratories (called *cristalizaderos*) during the whole year 2008. For some of them, relevant information is missing and leaves us with a workable sample containing 328 laboratories.

We acknowledge the fact that these laboratories do not constitute a random sample of the entire population of laboratories active in Colombia in 2008. More specifically, the representativity of the sample cannot be checked because information concerning the distribution of the population of laboratories in Colombia is not available. Nevertheless, we propose a comparison of the distribution of the laboratories contained in our sample by the department where the laboratory was operating and the corresponding figures contained in the coca cultivation survey 2008 published by the UNODC and containing information on illegal laboratories destroyed as provided by the national narcotics office of the Colombian government ([UNODC, 2009b](#)).

The comparison is presented in Table 1. Notably, the two distributions are very similar and in this sense our sample resembles pretty closer the distribution of seized laboratories provided by the Colombian government.

Table 2 presents the distribution of laboratories seized by region, month of seizure and counteracting force. The largest part of the laboratories destroyed

were in the regions of Pacific and Central. As for the month of seizure, the sample is well distributed across all the months of the year thus supporting the idea that the manufacturing of cocaine is a business going on all around the year. Indeed, coca bushes can be grown and harvested year-round with an average number of harvests of four per year (Mejia and Posada, 2008). Finally, the army is the counteracting body most active in the destruction of laboratories with a 52% of overall laboratories destroyed.

As far as the average potential production of cocaine of the laboratories destroyed is concerned, our figures do not show stark differences between the different counteracting bodies and the different months of the year. On the contrary, figure 2 reports the distribution by region and highlights that the regions of Meta-Guaviare, Pacific and Putumayo-Caqueta are those characterised by highly productive laboratories compared to other regions.

Econometric model

Since our aim is to provide an estimate of the amount of cocaine hydrochloride produced in Colombia during the year 2008, it is important to understand that the production process of cocaine hydrochloride from coca leaves is a chemical process that is typically performed in three steps:

1. Coca leaves are transformed into coca paste. The cocaine content of the leaves (dampened with lime water or other alkali) is extracted by means of kerosene, then with sulphuric acid a liquid solution of cocaine sulphate is produced. From this solution cocaine base precipitates (coca paste) with lime. Since this stage requires minimum levels of skills and investment (the few chemicals needed are easy to find and cheap), growers usually are able to produce coca paste, although, obviously, some farmers simply sell the coca leaves.
2. Coca paste is transformed into cocaine base. Coca paste is cleaned from impurities by means of sulphuric acid and potassium permanganate. Again alkali are needed to precipitate the free base, which is then dissolved in some types of solvents. Although this step requires additional skills and investment, many coca growers also perform it.
3. Cocaine base is refined into cocaine hydrochloride. This is done by adding concentrated hydrochloric acid that causes cocaine hydrochloride to sediment. This is a quite complex production process requiring

more skills, chemicals and investments. This process is typically performed in more sophisticated laboratories (*crystalizaderos*) usually run by organised crime groups, which are targeted by the counteracting units for their bigger scale of production, and this is the production step which our data refer to.

We put forward an econometric model aimed at analysing more deeply the production process of cocaine hydrochloride. In particular, we model the relationship between the different production factors (inputs) and the final product (output) in the production process of cocaine. As for the inputs of production, cocaine can be produced in several ways, and variations with regard to technique, reagents and ingredients are very likely. Nevertheless, three main factors of production can be singled out: labour, microwave ovens, chemical precursors (e.g. sulphuric acid, potassium permanganate, cocaine base, ether, hydrochloric acid, acetone and ethyl ether) (United-Nations, 1986; Mejia and Posada, 2008; EMCDDA, 2008).

In particular, our unit of reference is the single observed laboratory producing cocaine hydrochloride (*crystalizaderos*). By focusing on these laboratories as the final productive unit of pure cocaine, we are able to set aside concerns regarding several issues, such as the average quality of cocaine and the conversion of intermediate inputs of production deriving from other laboratories, which usually negatively affects the results of other works in the field (ONDCP, 2004; UNODC, 2009b).

We estimate two different models for the sample comprising full information for 166 laboratories. First, we estimate a Cobb-Douglas functional form (Cobb and Douglas, 1928) for the cocaine production function of the type:

$$LPotentialProduction_i = \beta_0 + \sum_{k=1}^4 \beta_k x_{ik} + \gamma^T Z_i + \epsilon_i \quad (1)$$

where the dependent variable $LPotentialProduction_i$ is the natural logarithm of the monthly potential production of cocaine of laboratory i . As for the independent variables, the term $\sum_{k=1}^4 \beta_k x_{ik}$ is composed of the four inputs of production. $LSolidPrecursors_i$ and $LLiquidPrecursors_i$ are, respectively, the natural logarithm of the amount of solid and liquid chemical precursors of laboratory i . $LEmployees_i$ is the natural logarithm of the number of people working in laboratory i . $LOvens_i$ is the natural logarithm of the number of microwave ovens used in laboratory i . Z_i is a vector of laboratory-specific control variables; and ϵ_i is the error term. In particular, we control

for the effects stemming from the month of seizure, the counteracting body carrying out the operation and geographical location by including a series of specific dummies.

Second, we estimate a translog functional form ([Christensen et al., 1971](#)):

$$LPotentialProduction_i = \beta_0 + \sum_{k=1}^4 \beta_k x_{ik} + \sum_{k=1}^4 \sum_{j=1}^4 \beta_{k,j} x_{ik} x_{ij} + \gamma^T Z_i + \epsilon_i$$

where the dependent variable, the main inputs of production and the controls are the same as in equation (1) while the term $\sum_{k=1}^4 \sum_{j=1}^4 \beta_{k,j} x_{ik} x_{ij}$ contains squared and interaction terms. When the latter term is equal to zero, the translog model is reduced to a Cobb-Douglas one.

Our dataset contains detailed information about the inputs of production for all the 328 laboratories contained in our sample, and 166 reports an expert evaluation of the potential production. Thus, an estimate of the potential production of these remaining laboratories is needed. To do that, we approximate the production process of cocaine for the 162 laboratories of which we lack an expert evaluation. We do so by estimating a ‘cocaine’ production function for the 166 laboratories of which we have an expert evaluation and use the model to estimate the monthly production potential for the remaining labs.

In this way, our dataset comprises 328 observations and, for each unit, information about several dimensions of the production process, such as the number of people involved, the quantity of cocaine leaves and cocaine hydrochloride; the quantity of both solid (e.g. potassium permanganate) and liquid (e.g. kerosene) chemical precursors; the number and the type of instruments seized (such as microwaves, scales, compressors, etc.); the geographical location of the laboratory being seized and, for 166 units, an expert evaluation — normally a chemist following the counteracting unit — of its monthly production potential.

Table 3 presents the descriptive statistics of the above mentioned inputs that are available for our sample of laboratories. As for chemical precursors, quantities are reported according to the nature of the precursor itself. If the precursor is a solid precursor, quantities are reported in metric tons while if the precursor is a liquid precursor, the quantity is reported in cubic meters. The conversion in the same unit of measurement would have needed information on the specific density of the precursor under consideration. This information is not available thus we decided to make two factors of production

out of one, thus chemical precursors are divided between solid and liquid precursors. The average laboratory employs 25 persons and processes cocaine using 4 microwave ovens, 1.75 metric tons of solid precursors and 7 m^3 of liquid precursors. Its potential production amounts to 3.5 metric tons per month.

As far as the correlation values among the independent regressors are concerned, apart from liquid and solid precursors, that are likely to be complementary inputs of production and for which the correlation value is the highest (0.56), the correlation across the independent variables is low, thus suggesting the absence of any relevant problems of multicollinearity. In order to check for its presence, we computed a variance inflation factor for the two variables. We found no indication of significant multi-collinearity amongst the two independent variables (i.e. the Variance Inflation Factor is 2.10 for solid precursors and 2.30 for liquid precursors, well below the threshold level of 5).

Results

This section is divided into three subsections. First, we present the results of the regression models proposed in the preceding section. Second, we provide details on the computation of our own estimate for the total production of cocaine in Colombia in 2008. Finally, we propose a set of validity tests aimed at increasing the robustness of our results.

Econometric results

The results of the econometric analysis are illustrated in Table 4. Column 1 presents the estimates of the translog regression model while column 2 reports the results for the Cobb-Douglas model.

Let us first consider the results for the translog specification. No coefficient is found to be significantly different from zero. We also report an F-test testing whether the term $\sum_{k=1}^4 \sum_{j=1}^4 \beta_{k,j} x_{ik} x_{ij}$ equals zero. Given that we cannot reject the null hypothesis at the 1% level, we find evidence that the correct model to test is reduced to the Cobb-Douglas one.

Let us now focus on the results of the Cobb-Douglas model (column 2 in Table 4). *LSolidPrecursors* exhibits positive coefficient, significant at the 5% confidence level. This means that the “solid precursors-cocaine” elastic-

ity is equal to 0.4. *LLiquidPrecursors* exhibits a positive and significant coefficient with an elasticity of 0.25. Finally *LEmployee* has a positive and significant coefficient with an elasticity of 0.86 while *LOven* is not significant. For the Cobb-Douglas specification we also test for the presence of linear homogeneity in the production of cocaine. The null hypothesis of linear homogeneity cannot be rejected thus suggesting the presence of constant returns to scale in the production of cocaine. This seems to indirectly confirm that laboratories have not become more productive thus contradicting one of the “hidden” reasons for labs increasing efficiency cited by the UNODC (UNODC, 2005).

Estimates

By means of the proposed models, we put forward an estimate of the potential production of cocaine as it has been accomplished by only those productive units whose seizure has been reported during the year 2008 by Colombian counteracting organisations. In this way we are able to provide an estimate for all the 328 laboratories, also the ones lacking an estimate provided by the seizing authorities.

The procedure to estimate the overall production of cocaine in Colombia in 2008 is the following. First, the monthly potential production of cocaine is estimated for each single laboratory contained in our sample using the results from the two different models. Second, the estimated production in 2008 of each laboratory is computed by multiplying the daily potential production and the number of days in which the laboratory has been in existence starting from the 1st of January 2008.

Before presenting the results of the estimation procedure we set forward some adjustments to the procedure described above. It is worth stressing that all of the mentioned statements are likely to provide downward estimates and, for this reason, the final estimate of cocaine production for the year 2008 that we propose here is highly conservative in nature. Moreover, we propose two different scenarios relating to the total amount of cocaine produced in Colombia during the year 2008: (i) 1.96 standard errors lower than the point estimate ($\hat{y} - 1.96 * se$) and (ii) equivalent to the point estimate (\hat{y}). Third, it must be recalled that we take into consideration only those laboratories that have been seized and reported by Colombian authorities. It is the case that not all of the productive laboratories are seized by the the counteracting organisations, otherwise Colombia will produce no cocaine at all, while it

is still internationally recognised as the major producer of cocaine in the world. Fourth, if more than one laboratory is found to be seized in the same locality, then we assume that any additional laboratories to the first one actually constitutes a replacement of the seized one and, for this reason, it contributes to the total estimate only for the additional days.

Given these conservative assumptions, the final numbers of our estimation exercise (see Table 5) are quite interesting and, we believe, shed new light on the way in which cocaine supply should be conceived by the relevant literature.

Even in the most conservative scenario, that is when we subtract 1.96 times the standard error of the estimator to the point estimate, production results to be 935 metric tons, a much higher value than the amount of cocaine reported by the UNODC and the ONDCP reports (UNODC, 2009b; ONDCP, 2011), which amount respectively to 450 and 295 metric tons. The estimate at the point estimate obtains the extremely large result of 2768 metric tons.

Validity tests

The final estimate provided in our work may be subject to several biases. For instance, depending on the nature of the missing values in the potential production of cocaine the coefficients estimated in our model can be biased. When data are missing completely at random (MCAR) no bias in the coefficient would result from ignoring incomplete observations and thus the estimate of missing values from the model would be correct. Thus, finding support for a missing data mechanism of the MCAR type would strengthen the reliability of our estimates. We propose two ways to check the credibility of our assumption that the missing data mechanism contained in our data is actually MCAR (see A for the details of the two procedures) and in both cases we find support for our claim that the missing data mechanism is completely random and adds no bias to our estimates.

In addition, we make some assumptions about the functional form of the production function, in particular we assumed that a linear relationship links the potential production of cocaine to the relevant inputs of production. Indeed, Cobb-Douglas and translog production function models specify in advance the form of the production function. While in a standard setting these forms can be implemented reliably, we cannot assume the same in an unusual setting such as the production of cocaine. For this reason, as a robustness check we implement a nonparametric regression approach (see B for

details). The estimate of cocaine production in 2008 following this alternative model specification, is still much higher than the estimates provided by official agencies (UNODC and ONDCP).

Third, the presence of severe outliers can actually bias our estimates. The bias may arise from the presence of few highly productive units in our data as evidenced by severe outliers. To cope with this limitation, we estimate a quantile regression model (see C for details) but, even in this case, estimates of total cocaine production for 2008 are still higher than official estimates.

Finally, the computation of the final estimate on the total production of cocaine in Colombia in 2008 is based on the strong assumption that all laboratories were in place at 1st January 2008. The average life length of a laboratory contained in our sample is 99 days. As further robustness check, we calculate the estimated total production for lower levels of average life length of the laboratories under consideration. By doing so, we show that, in order to reach an amount of cocaine production comparable to that from official agencies, the average life length of the laboratories contained in our sample should be reduced to 17 days. Unfortunately, there is no available information on the real average life length of a laboratory producing cocaine in Colombia but it looks reasonable to assume that it cannot be that short. Building a secret laboratory in the Colombian jungle is a financial investment which takes time and resources. Thus we expect crime groups to do that for an expected return lasting, on average, at least more than 1 or 2 months. If this is the case, then our estimate of the total production of cocaine in Colombia in 2008 increases to more than 1,000 metric tons, far above the one provided by UNODC and ONDCP.

Conclusion

In this paper, we put forward the idea that in cocaine-related studies the reliability of empirical data available to researchers is crucial. Although we witness a widespread interest in developing very visible policies heavily biased towards contrasting drugs supply, the quality of the data produced is not always following the trend. In an attempt to improve the reliability of data collection, we put forward a unique dataset covering seizures of laboratories for hydrochloride cocaine production in Colombia for the whole year 2008. By means of this dataset, we compute a production level of 935 metric tons in its lower boundary, which is more than twice than the official UNODC

estimates. In our calculations we were keen to use whenever possible the lowest boundaries of our estimates, in order not to inflate the final result. A set of validity tests has been performed, in order to confirm that our methodology is robust to a wide series of possible biases.

Our analysis shows that, on the grounds of a micro based dataset built through an alternative methodology to those commonly used, it is possible to contribute a meaningful analysis of such a complex phenomenon by overcoming some of the limits of official statistics. We thus furnish a tool of analysis that, if used together with the standard ones, could prove to be effective in complementing the information derived from the more usual statistical tools. Indeed, our effort is aimed at complementing, rather than challenging, the standard methodologies, in order to provide a more realistic and reliable view of such a complex phenomenon.

One last issue must be considered, related to the possible overestimation that the expert could be doing in their estimates of the potential production of the seized laboratory. Indeed, it might seem that counteracting forces have an incentive to overreport the results of seizing activity in order to show off. This also seems to be in line with recent findings obtained by [Matthew-Simmons et al. \(2011\)](#) who show that enforcement data are not always indicative of total supply, since there is a widely diffused effect of overestimation of seizures. At this regards, however, our data do not suffer from overestimation in their reporting, as they do not come from media reports, but directly from the counteracting units report, which is not per sé a cause of distortion.

However, more serious distortions might arise from other sources such as strategic or illegal behaviour. Indeed, if we suppose a strategic use of data on seizures, it might seem that a certain level of exaggeration in reporting the quantity seized could be expected, in order to show that counteracting units in general are effective in fighting against traffickers. However, by the same token the expert estimation of the production capacity of the laboratories should be kept lower than its true value (and not higher) since it seems reasonable that to act strategically counteracting forces should overestimate the quantity seized and underestimate potential production, in order to maximise their efficacy in counteracting illegal activities: “we seize a lot and thus potential production goes down”.

Another source of distortion may come from illegal behaviour. If we assume that part of the amount of cocaine seized is stolen by some members of the counteracting unit then data on both the quantity of cocaine seized and

potential production will be reduced in order to hide the stolen quantities.

Hence, in both cases, although we expect our data to be somehow biased, our expectations are that potential production could be even further underestimated.

Finally, our dataset seems to be immune from distortion in over reporting, as data are stationary (statistical tests are reported in [Leoncini and Rentocchini \(2010\)](#)), with no structural break suggesting the existence of such a phenomenon (apart from the October–December break, which is symmetrical and is probably due to the substitution of high ranks in the Army).

Future works should try to address all the points raised above to extend and complement our results. In spite of these limitations, we believe that our proposed methodology could prove to be an effective tool to parallel official statistics, as its main characteristics lend themselves to a wide diffusion of its usage. Moreover, the data on production reveal that a further effort is needed from the official agencies side in order to help practitioners to have increasingly reliable data on such an important phenomenon.

References

- Cameron, A., Trivedi, P., 2005. *Microeconometrics: methods and applications*. Cambridge University Press.
- Caulkins, J., Reuter, P., Taylor, L.J., 2006. Can supply restrictions lower price? violence, drug dealing and positional advantage. *Contributions to Economic Analysis & Policy* 5, 1–18.
- Christensen, L., Jorgenson, D., Lau, L., 1971. Conjugate duality and the transcendental logarithmic production function. *Econometrica* 39, 255–256.
- Cobb, C., Douglas, P., 1928. A theory of production. *The American Economic Review* 18, 139–165.
- Costa Storti, C., De Grauwe, P., 2009. The cocaine and heroin markets in the era of globalisation and drug reduction policies. *Int J Drug Policy* 20, 488–96.
- Costa-Storti, C., De Grauwe, P., 2009. Globalization and the price decline of illicit drugs. *Int J Drug Policy* 20, 48–61.

- EMCDDA, 2008. Monitoring the Supply of Cocaine in Europe. Technical Report. European Monitoring Centre for Drugs and Drug Addiction.
- Fries, A., Anthony, R.W., Cseko Jr., A., Gaither, C.C., Schulman, E., 2008. Technical Report for the Price and Purity of Illicit Drugs: 1981-2007. Technical Report. Institute For Defense Analyses.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica: journal of the Econometric Society* 46, 33–50.
- Leoncini, L., Rentocchini, F., 2010. Counteracting cocaine production. An analysis based on a novel dataset. Working Papers 693. Dipartimento Scienze Economiche, Universita' di Bologna.
- Little, R., 1988. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association* 83, 1198–1202.
- Matthew-Simmons, F., Shanahan, M., Ritter, A., 2011. Reported value of cannabis seizures in australian newspapers: Are they accurate? *Drug and alcohol review* 30, 21.
- Mejia, D., Posada, C.E., 2008. Cocaine production and trafficking: what do we know? Working Paper Series 4618. World Bank Policy Research.
- Nell, J., 1994. The dynamics of the drug market. *Challenge* 37, 13–21.
- ONDCP, 2004. The Price and Purity of Illicit Drugs: 1981 Through the Second Quarter of 2003. Technical Report Publication Number NCJ 207768. Washington DC: Executive Office of the President.
- ONDCP, 2011. 2011 International Narcotics Control Strategy Report. Technical Report. White House Office of Drug Control Policy (ONDCP). Available at <http://www.state.gov/p/inl/rls/nrcrpt/2009/>.
- Reuter, P., 2001. The limits of supply-side drug control. *The Milken Institute Review of Economic Studies* , 14–23.
- Thoumi, F., 2002. Illegal drugs in colombia: from illegal economic boom to social crisis. *Annals of the American Academy of Political and Social Science* 582, 102–115.

Thoumi, F., 2005. The numbers game: let's all guess the size of the illegal drug industry. *Journal of Drug Issues* 35, 185–200.

United-Nations, 1986. Recommended methods for testing cocaine: Manual for use by national narcotics laboratories. Technical Report. United Nations, Division of Narcotic Drugs.

UNODC, 2005. Coca cultivation surveys for Bolivia, Peru and Colombia. Technical Report. United Nations Office on Drugs and Crime.

UNODC, 2009a. 2009 World Drug Report. United Nations Office on Drugs and Crime.

UNODC, 2009b. Colombia coca cultivation survey. Technical Report. United Nations Office on Drug and Crime.

UNODC, 2010. 2010 World Drug Report. United Nations Office on Drugs and Crime.

Wood, S., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association* 99, 673–686.

Yatchew, A., 1998. Nonparametric regression techniques in economics. *Journal of Economic Literature* 36, 669–721.

Published on the International Journal of Drug Policy

Figures

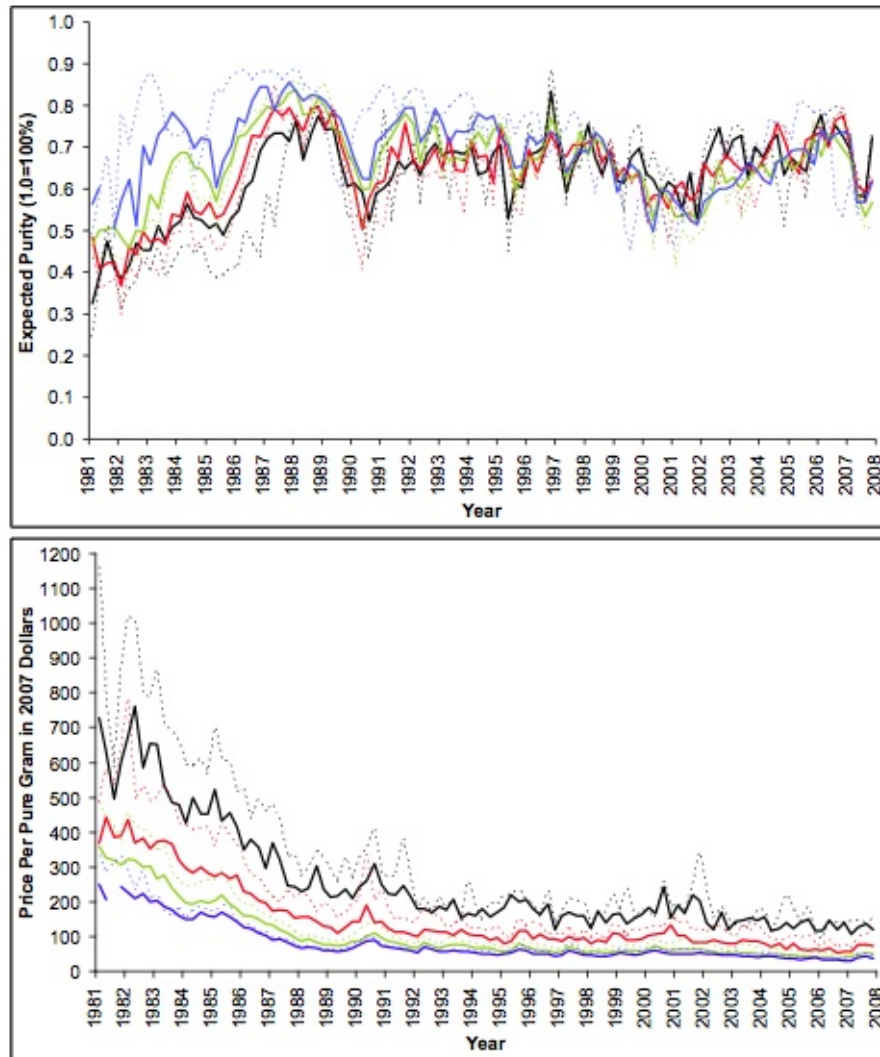


Figure 1: Price and purity of cocaine - Predicted price of one expected pure gram of powder cocaine (bottom) and expected purity of powder cocaine (top) compared to counterpart medians. Source: [Fries et al. \(2008\)](#)

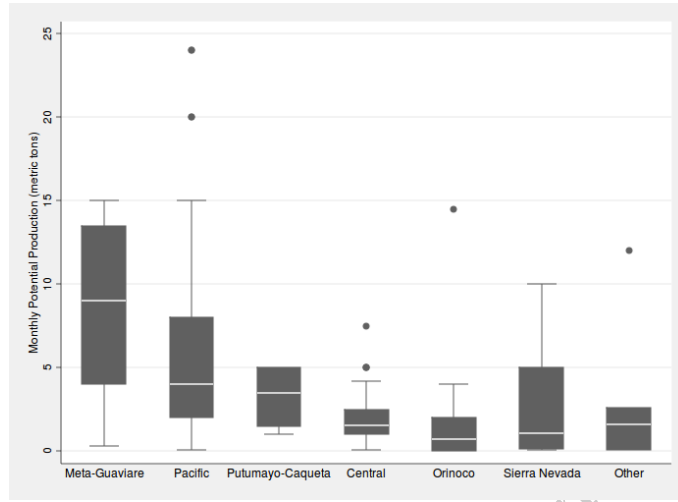


Figure 2: Monthly potential production of cocaine of seized laboratories by region of seizure

Published on the International Journal of Drug Policy

Tables

Table 1: Comparison of the number of laboratories seized by department in 2008: (a) our sample (b) UNODC report of year 2008 ([UNODC, 2009b](#))

| Department | (a) Our sample | (b) UNODC | Department | (a) Our sample | (b) UNODC |
|--------------|----------------------|--------------|-----------------|----------------------|--------------|
| Amazonas | - | - | Guainia | - | - |
| Antioquia | 32 | 37 | Guaviare | 2 | 2 |
| Arauca | 6 | 2 | Magdalena | 11 | 7 |
| Atlantico | 2 | 2 | Meta | 17 | 3 |
| Bolivar | 9 | 3 | Narino | 64 | 65 |
| Boyaca | 9 | 9 | N. de Santander | 49 | 27 |
| Caldas | 1 | 3 | Putumayo | 7 | 6 |
| Caqueta | 1 | - | Quindio | 1 | - |
| Casanare | 5 | - | Risaralda | - | 1 |
| Cauca | 61 | 42 | Santander | 4 | 14 |
| Cesar | 7 | 5 | Sucre | 1 | 1 |
| Choco | 5 | 5 | Tolima | - | 3 |
| Cordoba | 8 | 7 | Tunja | 2 | - |
| Cundinamarca | 3 | 6 | Valle | - | 11 |
| Guajira | 1 | - | Vichada | 20 | 6 |
| | | | Total | 328 | 267 |

Table 2: Distribution of seized laboratories by region, month of seizure and counteracting force

| | Freq. <i>n=328</i> | % Freq. | Freq. % Freq. <i>n=328</i> | | |
|------------------------------|-----------------------|---------|-------------------------------|----|------|
| Region | | | Month | | |
| Meta-Guaviare | 19 | 5.8 | Jan | 27 | 8.2 |
| Pacific | 130 | 39.6 | Feb | 33 | 10.1 |
| Putumayo-Caqueta | 8 | 2.4 | Mar | 24 | 7.3 |
| Central | 122 | 37.2 | Apr | 33 | 10.1 |
| Orinoco | 26 | 7.9 | May | 30 | 9.1 |
| Sierra Nevada | 12 | 3.7 | Jun | 18 | 5.5 |
| Other | 11 | 3.3 | Jul | 30 | 9.1 |
| Counteracting force | | | Aug | 23 | 7.0 |
| National Army | 171 | 52.1 | Sep | 35 | 10.7 |
| National Police | 125 | 38.1 | Oct | 44 | 13.4 |
| Other (Colombian Navy & DAS) | 32 | 9.8 | Nov | 22 | 6.7 |
| | | | Dec | 9 | 2.7 |

Table 3: Descriptive statistics for the variables used in the econometric models

| | Potential Production ¹ | Solid Precursors ² | Liquid Precursors ³ | Employees ⁴ | Ovens ⁴ |
|---------|--------------------------------------|----------------------------------|-----------------------------------|------------------------|--------------------|
| Mean | 3.54 | 1.75 | 6.99 | 24.69 | 4.43 |
| Median | 2 | 0.51 | 2.67 | 23 | 2 |
| Std.Dev | 4.39 | 4.29 | 12.35 | 14.4 | 6.79 |
| Min | 0.03 | 0 | 0 | 4 | 0 |
| Max | 24 | 41.5 | 110.91 | 137 | 50 |
| Obs | 166 | 328 | 328 | 328 | 328 |

¹metric tons per month; ²metric tons; ³cubic meters; ⁴n

Table 4: Results of ‘Cocaine’ production function regression models: factors influencing the monthly potential production of cocaine in Colombia for year 2008

| | | Translog | Cobb-Douglas |
|---------------|------------------------------|-------------------|---------------------|
| β_0 | Constant | -2.272 (3.647) | -2.282* (1.192) |
| β_1 | LSolidPrecursors | 1.460 (1.316) | 0.397** (0.157) |
| β_2 | LLiquidPrecursors | -0.920 (0.822) | 0.253*** (0.089) |
| β_3 | LEmployees | 1.123 (2.267) | 0.860*** (0.305) |
| β_4 | LOven | -0.785 (0.699) | -0.052 (0.088) |
| $\beta_{1,1}$ | LSolidPrecursors_sq | 0.040 (0.342) | |
| $\beta_{3,3}$ | LEmployees_sq | -0.216 (0.746) | |
| $\beta_{2,2}$ | LLiquidPrecursors_sq | 0.235 (0.164) | |
| $\beta_{4,4}$ | LOven_sq | 0.434* (0.255) | |
| $\beta_{1,3}$ | LSolidPXLEmpl | -0.280 (0.437) | |
| $\beta_{1,4}$ | LSolidPXLOven | -0.048 (0.143) | |
| $\beta_{2,1}$ | LLiquidPXLSolidP | -0.056 (0.136) | |
| $\beta_{2,3}$ | LLiquidPXLEmpl | 0.275 (0.271) | |
| $\beta_{2,4}$ | LLiquidPXLOven | -0.105 (0.100) | |
| $\beta_{3,4}$ | LEmplXLOven | 0.117 (0.236) | |
| | Regional controls | Included | Included |
| | Time controls | Included | Included |
| | Count. force controls | Included | Included |
| | F-test of Cobb-Douglas model | 0.25(1, 132) | |
| | F-test of linear homogeneity | | 2.16(1, 142) |
| | R^2 | 0.550 | 0.512 |
| | F-test of joint significance | 4.883(33) | 9.521(23) |
| | N | 166 | 166 |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are in parentheses.

Table 5: Comparison of overall estimates of cocaine production (in metric tons) in Colombia for year 2008: (a) own estimate, (b) UNODC official estimate ([UNODC, 2009b](#)), (c) ONDCP official estimate ([ONDCP, 2011](#))

| | (a) | (b) | (c) |
|---|------------|-------|-------|
| | Our sample | UNODC | ONDCP |
| Total production (lowest boundary level) ¹ | 935 | - | - |
| Total production (point estimate) ² | 2768 | - | - |
| Total production | - | 450 | 295 |

¹Result of the estimate produced with the Cobb-Douglas regression model keeping the estimates at the lower boundary level ($\hat{y} - 1.96 * se$)

²Result of the point estimate produced with the Cobb-Douglas regression model (\hat{y})

A Missing data

To account for the possibility of missing data, we hypothesise that the mechanism contained in our data is actually Missing Completely At Random (MCAR) (Cameron and Trivedi, 2005). First, we propose the results of the following logit regression model:

$$\text{Logit}[DMiss_i|x_{ik}] = \beta_0 + \sum_{k=1}^{10} \beta_k x_{ik} + \gamma^T Z_i$$

where $DMiss_i$ is a dichotomous variable taking value 1 if for that particular laboratory monthly potential production of cocaine was not provided by the chemist following the counteracting unit and 0 otherwise. $\sum_{k=1}^{10} x_{ik}$ are a set of covariates synthetizing the information contained in our sample of laboratories. Apart from the standard covariates used in the cocaine production function contained in the text, we included other variables (such as the amount of cocaine base, the number of hectares used for coca production, the amount of processed coca leaves and those before process etc.). Table 6 contains the results. Apart from the constant term, no other coefficient is significant at the usual significance levels thus supporting the claim that our missing data mechanism is MCAR.

A more formal procedure has been carried out to provide further evidence in favour of the MCAR nature of our missing data mechanism. In particular, we calculate the test statistic proposed by Little (1988) as a test of MCAR for multivariate data which yields a $\chi^2(9) = 9.6$. Thus, we do not reject the null hypothesis of MCAR providing in this way further evidence on the robustness of our estimates.

Table 6: Results of Logistic Regression Analysis: factors influencing the probability of missing value for monthly potential production of cocaine in Colombia for year 2008

| | Logit |
|------------------------------|-------------------|
| Constant | 2.575* (1.562) |
| Liquid precursors | -0.056 (0.054) |
| Solid precursors | 0.034 (0.043) |
| Employees | -0.041 (0.037) |
| Ovens | 0.001 (0.023) |
| Hydrochloric acid | -0.001 (0.001) |
| Instruments | 0.002 (0.027) |
| Hectares | -0.004 (0.049) |
| Cocabase | 0.002 (0.003) |
| Coca leaves | 0.182 (0.174) |
| Processed Coca leaves | -0.190 (0.223) |
| Regional controls | Included |
| Time controls | Included |
| Counteracting force controls | Included |
| Log-likelihood | -183.121 |
| $\chi^2(29)$ | 68.762 |
| Mc Fadden's R^2 | 0.194 |
| N | 328 |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Robust standard errors are in parentheses.

B Functional Form

To account for the possible bias of our estimates arising from the form of the function linking monthly potential production of cocaine to the inputs, we estimate a nonparametric additive regression model (Yatchew, 1998) which is able to trace the dependence between cocaine potential production and the inputs of production without specifying in advance the function relating the response to the predictors. In particular, we estimate the following model:

$$LPotentialProduction_i = \beta_0 + \sum_{k=1}^4 m_k x_{ik} + \gamma^T Z_i + \epsilon_i$$

where $\hat{m}_k(\cdot)$ are smoothing splines for the relevant covariates of our model obtained by the minimisation of a penalised sum of squares:

$$SS^*(h) = \sum_{i=1}^n [y_i - m(x_i)]^2 + h \int_{x_{min}}^{x_{max}} [m''(x)]^2 dx$$

where h is a smoothing parameter arising from the minimisation of the mean-squared error of the fit via generalised cross-validation (Wood, 2004).

Tabel 7 reports the results of the additive nonparametric model where dummy regressors are included in a linear form as they act as controls. We compute the estimate of Colombian total cocaine production in 2008 following the procedure presented above. Also in this case, if we subtract 1.96 times the standard error to the point estimate we get a total amount of cocaine production in 2008 much higher than the one provided in official statistics by the UNODC or ONDCP (450 metric tons), thus supporting our claim that production of cocaine is actually under-estimated by official public agencies.

Table 7: Results of Additive Non Parametric Analysis: factors influencing the monthly potential production of cocaine in Colombia for year 2008

| | edf | Ref.df | F |
|------------------------------|-----|----------|----------|
| LSolidPrecursors | 1 | 1 | 8.433*** |
| LEmployees | 1 | 1 | 5.262** |
| LOvens | 2.1 | 2.1 | 1.92 |
| LLiquidPrecursors | 7.6 | 7.6 | 3.320*** |
| Regional controls | | Included | |
| Time controls | | Included | |
| Counteracting force controls | | Included | |
| N | | 166 | |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C Outliers

Another bias may arise from the presence of few highly productive units in our data. In particular, the distribution in the monthly level of potential production of cocaine is highly skewed to the right, with a high number of abnormally low productive units, and with a fat tail to the right due to a non negligible number of highly productive laboratories. Moreover, for many observations our dependent variable presents strong departure from the mean value, thus calling for an evaluation at different points of the conditional distribution. To cope with these problems, we estimate a quantile regression model (Koenker and Bassett, 1978) of the following type:

$$Q_q(LP_{\text{PotentialProduction}}_i|x_i) = \beta_0 + \sum_{k=1}^4 \beta_k x_{ik} + \gamma^T Z_i + F_{u_i}^{-1}(q)$$

where $Q_q(LP_{\text{PotentialProduction}}_i|x_i)$ denotes the q th conditional quantile function of the potential production of cocaine for laboratory i given a set of covariates x_i . $F_q^{-1}(u_i)$ represents the inverse of the distribution function of error term u_i while the other terms are the same as in equation (1).

Table 8 shows the results of the quantile regression evaluated at five different quantiles (0.1, 0.25, 0.5, 0.75 and 0.9). Apart from the coefficient of liquid precursors, the significance and sign of other coefficients is in line with the results obtained with the Cobb-Douglas specification. Output elasticities tend to increase, as “solid precursors-cocaine” elasticity is now 0.7, while “labour-cocaine” elasticity raises to 1.2. Nevertheless, estimates of total cocaine production for 2008 is still higher than official estimates, even if we take into consideration the lowest quantile in terms of potential production (quantile 0.1 with an estimate of 999 mt for 2008). Indeed, to obtain an estimate close to that provided by the UNODC (i.e. 430 mt) we need to rely on the results from the quantile regression for the 0.25 quantile of laboratories and subtract one standard error from the point estimate.

Table 8: Results of Quantile Regression Analyses: factors influencing the monthly potential production of cocaine in Colombia for year 2008

| | Quantiles | | | | |
|-----------------------|---------------------|-------------------|---------------------|--------------------|---------------------|
| | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
| Constant | -7.047** (3.073) | -2.511 (2.705) | -3.815 (2.355) | -2.521 (1.801) | -3.937* (2.014) |
| LSolidPrecursors | 0.609*** (0.228) | 0.346 (0.213) | 0.681*** (0.236) | 0.500** (0.211) | 0.465** (0.220) |
| LLiquidPrecursors | 0.033 (0.182) | 0.224 (0.160) | 0.123 (0.123) | 0.066 (0.103) | 0.001 (0.111) |
| LEmployees | 2.029** (0.830) | 0.744 (0.699) | 1.182* (0.608) | 1.169** (0.475) | 1.555*** (0.532) |
| LOven | 0.027 (0.160) | -0.095 (0.143) | -0.146 (0.110) | -0.136 (0.094) | -0.036 (0.120) |
| Regions | Incl. | Incl. | Incl. | Incl. | Incl. |
| Time | Incl. | Incl. | Incl. | Incl. | Incl. |
| Count. force controls | Incl. | Incl. | Incl. | Incl. | Incl. |
| Pseudo R^2 | 0.56 | 0.4 | 0.28 | 0.34 | 0.39 |
| N | 166 | 166 | 166 | 166 | 166 |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors, obtained via 500 bootstrap replications, are in parentheses.