UNIVERSITA' DEGLI STUDI DI BOLOGNA

DIPARTIMENTO DI SCIENZE ECONOMICHE

Counteracting cocaine production
An analysis based on a novel dataset

Riccardo Leoncini
Francesco Rentocchini

Quaderni - Working Papers DSE N° 693

Counteracting cocaine production
An analysis based on a novel dataset*

Riccardo Leoncini†
Francesco Rentocchini‡

February 3, 2010

Abstract

The debate about the effectiveness of the counteracting policies against the supply of drugs, in particular of cocaine, is very lively and intense. Indeed, since many opinions are based on certain measures rather than others, the construction of reliable indicators is one of the preconditions for a correct and concerted assessment of drug supply. The lack of reliable data on drug provision derives, on the one side, from the objective difficulties encountered in assessing the quantitative elements of drug production and drug trafficking due to its illegal nature, and, on the other side, from the lack of a standard methodological approach to the issue. This paper tries to contribute to the topic by proposing a new dataset, based on a completely new approach to the problem of measuring drug supply. We put forward a unique dataset covering cocaine related seizures in Colombia for the whole of year 2008. Data have been collected on a daily basis from the websites of the main organizations fighting against drug traffickers (Army, Air Force, National Police, Departamento Administrativo de Seguridad, Armada Nacional, Fiscalia), detailing each single seizure of laboratories for the production of both basic paste and cocaine hydrochloride. By means of this dataset, we offer some accounts of the main numbers on drug supply and on drug seizures, suggesting some policy options, and arriving to an estimate of cocaine production.

JEL: K42, C81, D78
Keywords: war on drugs, cocaine production, drug control policy, Latin America, Colombia

*A preliminary version of this paper was presented to the Conference "Cento anni di controllo internazionale sulle droghe", Forlì, 21-22 April 2009 and to the ESSD Annual Conference, Belfast, 24-26 September 2009. The authors are grateful to the participants for comments and suggestions. Sandro Donati was instrumental for our decision to commit ourselves to the difficult task of elaborating these new and, hopefully, provocative ideas: we thank his intelligence, patience and enthusiasm. The usual caveats apply.

†Dpt of Economics, University of Bologna (I) and Ceris–CNR, Milan (I), riccardo.leoncini@unibo.it

‡Dpt of Economics, University of Trento (I), francesco.rentocchini@economia.unitn.it

1 Introduction

The last fifteen years have witnessed an increasing interest in drug–related issues. At both micro and macro level very important topics have been put on the agenda. They are related to several topics, ranging from health to ecology, from financial to industrial economics, from socio–political implications to terrorism.

In particular, the increased interest was mainly driven by a peculiar attitude, that has been developed by almost all the countries' government in the world, aimed at developing very visible policies heavily biased towards contrasting the supply–side of this phenomenon. The emphasis of all the counteracting policies developed by the vast majority of countries with regards to drugs is on an uncompromising prohibitionist approach, which has informed all the action plans aimed at eradicating drug trafficking. The reduction in supply is therefore seen as the appropriate answer to fight against drugs and drug–related offenses. The emphasis on repression and punishment is witnessed by increasing budgets for counteracting forces, by heightening incarceration, by new measures to counteract drug trafficking, etc (see, for instance, Mejia, 2008). Yet, in spite of this acute interest, data produced are so poor to render the attempts at understanding these phenomena meaningless in many cases, and at least questionable in the vast majority.

The lack of reliable data[1] in general, and on supply in particular is due to a set of both general and specific elements. The general elements are relative to the illegal nature of the objects of analysis, to the need to supply heterogeneous kind of data for many different final users (e.g. academicians, politicians, consultants, journalists, army) with many different aims, to the absence of established, consistent and mutually recognized methodologies.

Each of these factors produces different distortions with several implications. For instance, the largely illegal market for drugs implies that at each step some assumptions on latent variables have to be made,[2] and each assumption, far from being neutral, comes at a price. The price is to add a multiplicative bias for each of the steps followed that can result 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.

The specific elements[3] are instead related to the particular methodology with which the supply is determined. This leads to a diffused sense of unreliability of the data on drug production, even when they come from important agencies

[1] See, for instance, the introduction (Van Ours and Pudney, 2006) of the special issue of *De Economist*, where this problem is one of the themes addressed.

[2] For instance, there is not a proper market indicating quality or scarcity as function of price "only", there is not an agreed–upon quantity of hectares cultivated, there is not a set of established transformation coefficients for the quantity of leaf from an hectare of cultivated land, etc.

[3] For instance, the methodology adopted by the main data producers rely on aerial acquisition of cultivated land, from which data on tons of leaves per hectare are inferred, then through some transformation coefficients of leaves into base paste, and then of base paste into cocaine hydrochloride, an estimate of cocaine production is arrived at.

such as the United Nations Office on Drugs and Crime (UNODC) and the White House Office of National Drug Control Policy (ONDCP). Indeed, starting from early works, (Reuter and Greenfield, 2001; Thuomi, 2003), until a very late paper from Donati (2009), the data on cocaine production and seizures (as produced, for instance, by UNODC, 2009) have been severely questioned. To add further confusion to the picture, the data have been repeatedly corrected in order to cope with several inconsistencies, leading eventually to unbelievably high proportions of seizures compared to production (see for instance, the thorough account given in Donati (2009) about this latter issue).

The mix of objective and subjective difficulties surrounding data production has prompted an effort to try to assemble a completely new dataset, based on a completely new set of basic information. Indeed, our main objective is to build up a set of brand new indicators useful to gain deeper insight on the phenomenon of drug trafficking from a radically different viewpoint. This effort is aimed at complementing, rather than challenging, the more standard methodologies, in order to provide a more realistic and reliable view of such a complex phenomenon. With this effort in mind, the aim of this paper is thus to show the main characteristics and the main statistical properties of a unique dataset covering **daily** cocaine–related seizures in Colombia for the whole year 2008, carried out by **all** the counteracting organizations that operate on the field.

We will carry out some elaborations, with the aim to show how this database constitutes a completely new and original attempt at building a solid empirical ground to complement and enhance the analysis of cocaine production. In this way, we think we will provide reliable data to evaluate the impact of cocaine production and trafficking, and to elaborate effective counteracting policies. First of all, we will show the main characteristics of the dataset, in order to give a full picture of its most important statistical properties. Subsequently, we will show how such a dataset can be of help in offering some novel insights on the topic, in order to suggest how some of the results obtained therein can be useful for the purpose of counteracting policies. In so doing, we offer an instrument that, to our knowledge, has never been used before, and that could constitute an interesting and useful complement to more standard types of analysis.

The paper is organized as follows. In Section 2 the main problems linked to the evaluation of the cocaine supply will be presented and discussed. Section 3 will present the dataset and its main statistical features. Finally, Section 4 and Section 5 will show some econometric exercises aimed at showing the main results from a policy perspective. Conclusions will follow.

2 Limits to standard evaluations methods

It is now a commonly agreed fact that scholars do not agree on the reliability of the data utilized in the studies on drug–related issues.[4]

[4] See, for instance, Horowitz (2001) and the reply from Arkes, Pacula, Paddock, Caulkins, and Reuter (2008), to get a flavor of the the lively debate on the reliability of STRIDE data on acquisition of drugs by undercover agents and informants, in order to collect data on price

The problems with this kind of topic involve a huge number of issues related, first of all, to the inner multidisciplinarity of the research approaches to drug–related issues (and in particular, for this paper's sake, to the supply side of the phenomenon). Just to briefly scratch the surface of the problems involved, studies on drug–related issues come from research approaches as varied as ethnography, psychology, economics, criminology and law, all with their different methodologies (qualitative, quantitative, survey–based, behavioral research). To add further complication, it has to be considered that each of the above mentioned approaches can imply different elements related to both positive and normative criteria: this adds further implications related to stark differences in dealing with the main policy issues (such as, law enforcement, prohibitionism, harm reduction, prevention, trafficking counteracting, etc.) with a set of issues as their, so–to–say, corollaries (such as, school based prevention, war on drugs, money laundering, medical treatments, psychological rehabilitation, etc.).[5]

Therefore, as there are no standard ways to treat such a complex topic, it is impossible to take for granted elements that are elsewhere considered as exogenous parameters to the problem to be solved. Indeed, just to make as example, in production theory, economists usually can take for granted the engineering side of the production process, as the literature almost unanimously agrees on a set of standards to which firms must obey not to be overthrown from the market. In the case of our analysis, the degrees of freedom in production are so many to guarantee that the margins of efficiency are wide enough to allow for the survival of even really inefficient units of production. The reason lies in the impossibility for the final consumer to check objectively the quality and to find alternatives to the one he/she is supplied with. In this sense the black box of production is quite wide, and many different arrangements can produce (notionally) the same amount of final output.

Indeed, for instance, (i) prices are not global, but rather are fixed differently on different (local) markets (Wilson and Stevens, 2009); (ii) the price and the quality of the drugs sold vary wildly both geographically (Reuter and Greenfield, 2001) and along the distribution chain (ONDCP, 2004, ch. 3); (iii) supply (and demand as well) can be scattered and infrequent due to exogenous factors that can hardly be modeled, such as police, fights and so on (Levitt and Venkatesh, 2000, e.g.). All these elements boil down to the impossibility of analyzing this kind of market by means of the usual statistical tools. Indeed, since the data are invariably punctuated, there are two unintended consequences from using them: (i) data should better be treated "as a sample from a distribution of values rather than as observations of a single number" (Reuter and Greenfield, 2001, p. 170); (ii) data are usually presented as intervals, but the use of intervals makes it very hard to say something about the absolute dimension and, more

(in case of acquisition as opposed to seizures), quantities and purity.

[5]See Ritter (2006) for a thorough review of the main elements involved in evaluating the many different disciplinary approaches, and to the many caveats needed to produce a result, and to the many differences in the results obtained by the various fields of analysis.

importantly, even about the direction of the ultimate dynamics of such data.[6]

In this paper, we will focus our attention to the very long and lively debate on the data on cocaine supply and trafficking, which are officially produced by United Nations Office on Drugs and Crime (UNODC) and by the White House Office of National Drug Control Policy (ONDCP). Indeed, as many researchers (Reuter and Greenfield, 2001; Mejia and Posada, 2008; Donati, 2009) have convincingly shown, data on drug trafficking and seizures show a low degree of precision, which, in turn, characterizes the whole set of data on these phenomena that scholars are used to work with.

Therefore, just to make some examples, data on production are based mainly on aerial recognition that determines an estimation of hectares of bush cultivation, for which it is impossible to apply the usual statistical tools to determine the appropriate sample size. This, in turn, is translated into production by a conversion factor for the amount of plant for hectare, which is very difficult to obtain and has to be estimated. Still, to have data on cocaine production, sun–dried coca leaves have to be converted into coca paste with, at least, two other elements of discretionality/distortion (conversion rate and purity). Again, coca paste is transformed into coca base with variable degrees of quality conversion ratios, and especially chemical precursors. Finally, the same patterns is repeated for transformation into final product (cocaine hydrochloride). In between, with different degrees, cocaine base paste can be consumed[7] and/or transformed into crack.

In spite of these limitations, interesting models, focused on both demand and supply, have been produced to describe the likely behavior of this market, to quantitatively evaluate some important demand variables, such as price elasticity (e.g. Mejia, 2008), the data on production and trafficking (e.g. Mejia and Posada, 2008) and the impact of globalization on cocaine market (e.g. Costa Storti and De Grauwe, 2009), and to quantitatively evaluate the effectiveness of various policy options against drug production and trafficking, such as, for example, the interdiction policies (e.g. Caulkins, Reuter, Iguchi, and Chiesa, 2005; Fowler, 1996), coca eradication and the consequent decisions about how and what cultivate in Colombia (e.g. Moreno-Sanchez, Kraybill, and Thompson, 2003; Ibanez, 2006; Tabares and Rosales, 2005), the effectiveness of prevention and treatment policies (e.g. Mejia, 2008), making the illegal activities riskier and legalizing previously illegal activities (e.g. Chumacero, 2008), tougher enforcement against sellers (e.g. Caulkins, Reuter, and Taylor, 2006).

It is also possible to show that by taking into account the true dynamic nature of such a market, the standard results of comparative static economic analysis (Caulkins, Reuter, and Taylor, 2006) can be both easily overturned and made more adherent to the reality of the empirical observations. In particular, Nell (1994) shows that, by taking into account a truly dynamical approach, irreversibilities arise such that the development process of the counteracting

[6] Just to give an example, Thuomi (2003) reports estimates of Colombian net cocaine revenues ranging from US$ 194 to US$ 2.625 million in 1994!

[7] With different names, different usage, and different cocaine contents. See, for instance, VVAA (2006) for the case of Latin America.

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

Because of these many reasons, we have decided to try another way to deal with cocaine production and the counteracting activity, based on a completely different methodology. For this reason, we aim at furnishing to the researchers another tool, that can be used jointly with the more standard ones. With this purpose in mind, we have collected in a database all the seizures performed by all the counteracting units working on the field in Colombia.

In the following paragraph, we will discuss the main statistical characteristics of the dataset, in order to show that its main statistical properties are amenable to provide us with a useful tool to make some positive and some normative analysis.

3 The dataset

The data we use in this paper have been collected on a daily basis, through the press releases available on the websites, of all the main organizations – Army, Air Force, National Police, Departamento Administrativo de Seguridad (DAS), Armada Nacional, Fiscalia – 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 collect data detailing each single seizure of laboratories, for the production of both basic paste and cocaine hydrochloride. For each seizure a press release is usually published by the organization that accomplished the operation. Two types of seizures are reported: (i) laboratories where base paste of cocaine is produced, (ii) laboratories where the cocaine hydrochloride is produced (called "cristalizaderos"). In the press releases several other data are reported as well: the geographical location where the seizure took place; the number of people involved; the quantitative of either cocaine leaves, cocaine base paste or cocaine hydrochloride; the quantitative 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 seizures, an expert evaluation — normally a chemist following the counteracting unit — of the production potential of that particular laboratory is supplied.

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 seizures 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, to our dataset, severe outliers. 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 organizations, etc.) which allowed us to split the data across the days of the mission. Finally, in order to avoid replications that might have gone unnoticed, in case of seizures occurring during the same day in the same region, we opted to include only one.

After the cleaning–up, the dataset comprises 1.900 counteracting operations carried out during the period between the 1st of January and the 31st of December 2008. In this time span, the 1.900 operations led to the seizures of 2.259 laboratories of cocaine base paste and 333 cristalizaderos, for a total of 2.592 seizures during the whole year 2008.

It is worth noticing that such a dataset can be analyzed in two different ways: (i) as a time series detailing the number of laboratories and cristalizaderos seized daily in the period between the 1st of January and the 31st of December 2008; (ii) as two cross-sectional datasets containing detailed information on each single unit producing either base paste (laboratories) or cocaine hydrochloride (cristalizaderos). The following sections are thus dedicated to a detailed analysis of the dataset from these alternative viewpoints. In particular, Section 4 studies the behavior of time series, relating to seizures of both cristalizaderos and laboratories, by providing descriptive statistics on the series, checking for the presence of stationarity in the series and investigating the presence of structural breaks. In Section 5 we turn to a cross-sectional analysis and investigate a reliable production function of cocaine hydrochloride and provide an estimate of the total production of cocaine for year 2008.

4 Time series analysis

The current section mainly presents descriptive statistics relating to the dataset conceived in a time series fashion. In particular, we present some basic statistics and graphs which describe the two main series in a preliminary way. Moreover, we investigate the presence of stationarity for the series, conceived as a first step to model the underlying generating processes. After that, we put forward an Auto Distributive Lag (ADL) model aimed at explaining the number seizures for the cristalizaderos. Far from being able to provide a complete explanation of the dynamics of this phenomenon, the model nevertheless helps uncovering some interesting aspects in the series, namely the presence of structural breaks. We further investigate this issue by relying on rolling regressions.

In Table 1 we report the basic descriptive statistics. As shown, a total of 365 observations, with an average of 1 cristalizaderos and 6 laboratories per day, are found. The distributions of the two series under consideration present similar values for the mean and the median. In addition, both of them are right skewed (skewness > 0) and present fat tails (kurtosis > 3) thus indicating, in some cases, abnormal seizures in particular days of the year. Furthermore, the variability is the same for both series.

Table 1: Descriptive statistics

	Labs	Cristalizaderos
Mean	6.19	0.91
Median	5	1
Standard Deviation	5.72	1.11
Skewness	1.74	1.47
Kurtosis	7.54	5.26
Min	0	0
Max	35	6
N. Obs.	365	365

By plotting the cycle-plot for the total number of seizures, we have been able to depict behaviors which reveal interesting aspects concerning the way the work of the counteracting organizations take place. In particular, Figure 1 shows a clear day-of-the-week effect with low numbers of seizures on Sundays, probably due to vacation days, and high levels of seizures on Mondays and Tuesdays, probably because these are the reporting days for operations carried out during the week-end. This behavior was also confirmed by running a regression analysis using the days of the week as controls.

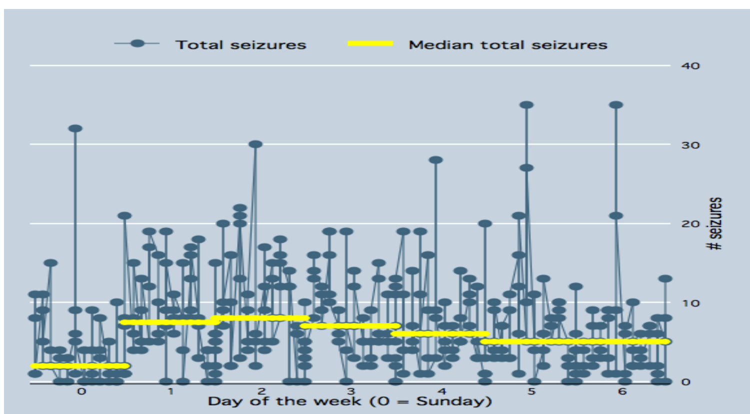

Figure 1: Total number of seizures by day of the week

In Table 2 we provide indications on level stationarity of the series concerning the number of daily seizures for laboratories of base paste, cocaine hydrochloride and the total number of seizures. In this way we have been able to correctly develop an econometric estimation without incurring in the problem of spurious regression where both estimators and test statistics are misleading. As reported in the Table, the series of seizures referring to cristalizaderos reject the null of presence of unit roots at the appropriate lag and, at the same time, accept the null of level stationarity. On the contrary, the series containing the total number of seizures per day, albeit not revealing the presence of unit roots at the appropriate lag, it nevertheless rejects the null of level stationarity. This is clearly driven by the behavior of seizures for the laboratories, as revealed by the results of the two tests for this series. As expected, laboratories are integrated of order one and, by differentiation, they turn to level stationarity (the same applies to the series on total seizures).

Table 2: Tests for the presence of unit roots and level stationarity in the series of seizures for total seizures, cristalizaderos and laboratories

Series	Level stationary*	No unit roots**
Seizures_tot	No	Yes
Cristalizaderos	Yes	Yes
Labs	No	Yes
D.Labs	Yes	Yes
D.Seiz_tot	Yes	Yes

* The null hypothesis of level stationarity is accepted at the 1% level according to the test proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992).
** It rejects the null hypothesis of the presence of unit roots according to the test proposed by Elliott, Rothenberg, and Stock (1996) at the appropriate lag according to the Modified AIC (Ng and Perron, 2001).

Being the series level stationary, we can now investigate some interesting issues concerning the extent of correlation between them. Indeed, we propose to model the number of daily seizures of cristalizaderos as a function of the lagged number of seized cristalizaderos plus the number of laboratories producing base paste. Far from being a reliable model for time-series forecasting given the lack of important explanatory variables likely to explain the dynamics, it nevertheless will allow us to uncover some interesting aspects which might deserve further analysis. In particular, we put forward an Auto Distributive Lag model (ADL) aimed at explaining the amount of seizures of cristalizaderos by day as a function of the number of seizures for cristalizaderos and laboratories in the previous days. As expected, the model presents problems of auto-correlation which cannot be reduced by differencing (the null of no autocorrelation for the Breusch-Godfrey test is rejected at the 1% level). In particular, this is probably due to a problem of endogeneity arising from an omitted variable bias, that is, important factors explaining the counteracting activity of cocaine fails to be addressed by the current model. In general, the model returns unreliable estimates for the parameters of interest but, at the same time, it can be adopted to obtain some indications for further analysis. In particular, we are interested in the presence of sensible changes in the amount of seizures during the period under consideration. Table 3 presents results for the estimation of the following model:

$$Crist_t = \alpha_0 + \alpha_1 Crist_{t-1} + \beta_0 \Delta Lab_t + \sum_{i=2}^{12} \lambda_i Month_i$$

where $Crist$ stands for the daily number of seizures of cristalizaderos, ΔLab_t is the first difference in the series of laboratories and $Month$ is a dummy variable taking value 1 for a particular month of the year and zero for all the other months. In this way, we are able to check whether in particular months of the year a sudden change in the amount of seizures of cristalizaderos is taking place. Results clearly show that in October and December there has been an opposite shock in the number of seizures of cristalizaderos.

Table 3: ADL(1,1) model controlling for month of seizure

$Crist_t$	Estimate	St. Error	Significance
Constant	0.828	0.201	***
$Crist_{t-1}$	0.029	0.052	
Δ Lab	0.036	0.010	***
February	0.286	0.279	
March	-0.044	0.274	
April	0.292	0.276	
May	0.124	0.274	
June	-0.255	0.277	
July	0.117	0.274	
August	-0.112	0.274	
September	0.358	0.277	
October	0.558	0.275	*
November	-0.107	0.276	
December	-0.537	0.278	*
N. Obs.	364		
R^2	0.105		

* p<0.05, ** p<0.01, *** p<0.001

This result is also confirmed by checking for parameters instability via rolling analysis. This is accomplished by saving the parameters of the model as it is estimated in a window rolling along the time axis thus producing estimates for overlapping time frames (the so–called rolling regressions or moving–window regressions). In our case, we set the window to a width of 50 days and, in this way, for each day starting from the 51st we compute an estimate based on the previous 50 days. Overall, we are able to depict the dynamics of the coefficients in the ADL model, thus observing its behavior. Clearly, if a series receives an exogenous negative shock, it will be characterized by a sudden drop in the parameter and vice versa. In this way, we can provide an accurate description of the evolution through time of the time series and of the existence of eventual structural breaks. Figure 2 confirms the preliminary results obtained in the estimation of the ADL model. Indeed, the values of the coefficients rise in October and fall in December, together with an additional structural break in August.

Moreover, during the summer the predicting power of seizures in the labs over those of cristalizaderos are found to be reversed (Figure 3), suggesting that the reasons for the break should be found elsewhere.

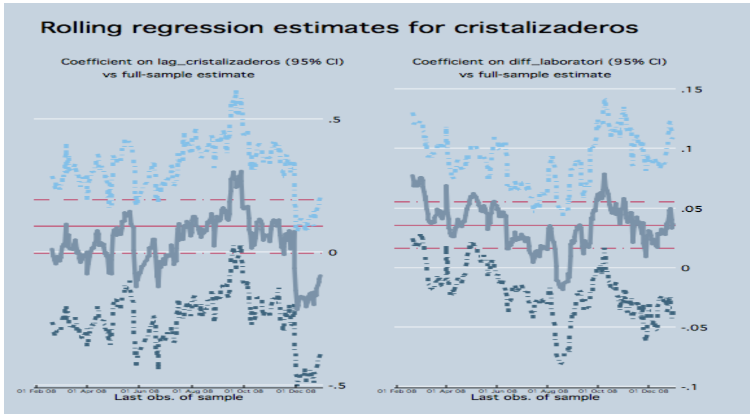

Figure 2: Parameter stability by rolling regression analysis

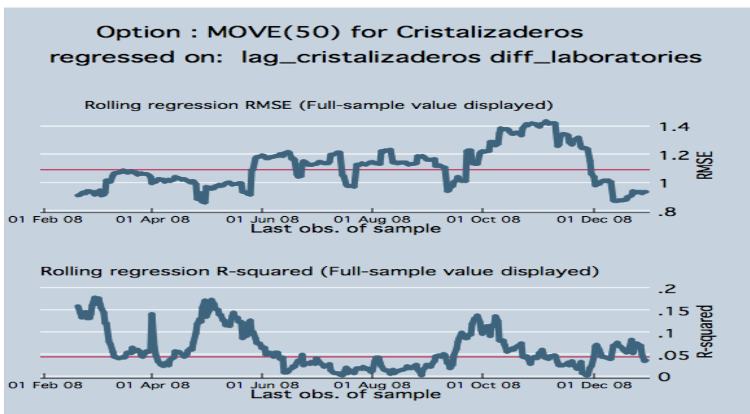

Figure 3: Stability in the predicting power of the ADL model

5 Cross–section analysis

Cross–section analysis has been carried out having as unit of reference the single observed laboratory producing cocaine hydrochloride (cristalizaderos). By focusing on cristalizaderos 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, 2009). In this way, our dataset comprises 328 observations and, for each unit, information about several dimensions of cristalizaderos' production process, such as the number of people involved, the quantitative of cocaine leaves and cocaine hydrochloride; the quantitative 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 the production potential of that particular laboratory.

The analysis carried out in this Section is divided into three main steps. First, given the high variability characterizing many features of the units of production, observations have been grouped in different clusters in order to uncover the underpinning patterns. Second, we provide results of a quantile regression model aimed at scratching the surface of the production function of cocaine. Finally, building upon results obtained from the previous analysis, we provide a parsimonious estimate of production of cocaine hydrochloride for the year 2008 and compare it with official statistics from international organizations.

Table 4 shows general descriptive statistics of the relevant variables concerning the production process of cristalizaderos. As evidenced in the table, the amount of both solid and liquid chemical precursors, as well as the number of ovens and people therein working, are characterized by a high variability. The high extent of variability may hide important differences in the fundamental features of the production process corresponding to different abilities in the production of cocaine. For this reason, we first run a cluster analysis to identify groups of cristalizaderos with specific characteristics and to provide some clues concerning the provision of tailored counteracting policies.

Table 4: Descriptive statistics

	Potential production *tons per month*	Liquid precursors *gallons*	Solid precursors *tons*	Ovens *n*	People *n*
Mean	3.54	1845.3	1.75	3.71	11.11
Median	2	705	.505	1.5	0
Coeff Var	1.24	1.77	2.46	1.62	1.9
Std Dev	4.39	3263.74	4.29	6.02	21.15
Skewness	2.48	3.96	5.63	3.34	2.81
Kurtosis	9.83	24.77	42.16	19.36	13.22
N	166	328	328	328	328

Relevant information on the three main clusters[8] is reported in Table 5, which helps shedding light on the highly skewed distribution in the productivity of single cristalizaderos. As it can be easily seen, the three clusters strongly differ in size and some other relevant features. As for cluster 1, it contains a small number of extremely highly productive cristalizaderos (14) which make

[8]We tried different procedures for identifying clusters of cristalizaderos. As a preliminary step, we run hierarchical cluster analysis to identify the range in the optimal number of clusters. The different methods yield a number of clusters ranging from 2 to 6. We employ a partition cluster analysis by means of k-means clustering and we set the optimal clusters' number according to the Caliński and Harabasz (1974) pseudo-F statistic.

an intensive use of chemical precursors. Cluster 2 contains the highest number of cristalizaderos (248) with a lower–than–average productivity. Finally, cluster 3 contains 66 labour–intensive cristalizaderos with a productivity level close to the sample average.

Table 5: Results of cluster analysis

Cluster		Potential production *tons per month*	Liquid precursors *gall*	Solid precursors *ton*	Ovens *n*	People *n*
1	N	11	14	14	14	14
	Mean	11.01	12539.64	15.5	15.36	27.29
2	N	108	248	248	248	248
	Mean	2.38	1138.39	0.99	2.79	2.08
3	N	47	66	66	66	66
	Mean	4.44	2233.09	1.69	4.71	41.59
Total	N	166	328	328	328	328
	Mean	3.54	1845.3	1.75	3.71	11.11

It is interesting to analyze the three clusters keeping in mind, for instance, the possibility of tailoring policies against cocaine production. In particular, Tables 6, 7 and 8 show, for each cluster, the distribution of seizures by area, by the counteracting unit and by month of the year. Interesting results are highlighted in grey in the tables: (i) the most productive cristalizaderos, belonging to cluster 1, are seized in the area of Meta-Guaviare and Pacific while productive units belonging to the other two clusters are mainly located in Pacific and Central areas (Table 6); (ii) the Army is the counteracting unit seizing mostly cristalizaderos belonging to cluster 2 and 3, while the most productive cristalizaderos are seized equally among the different counteracting units (Table 7); (iii) seizures of the most productive cristalizaderos (cluster 1 and cluster 2) concentrate during the first months of the year while the least productive cristalizaderos are equally distributed along the year with a peak during the last months (Table 8).

As for the second step in the cross–sectional analysis of the collected data, we put forward an econometric model aimed at analyzing more deeply the productive process of cocaine hydrochloride. Our main interest is to understand, at least partially, the existence, direction and intensity of the relationship between the different production factors (inputs) and the final product (output) in the production process. The model tested can be represented by the following equation:

$$Q_q(PPC_i|x_i) = \alpha_1 + \beta_1 SP_i + \beta_2 LP_i + \beta_3 F_i + \beta_4 P_i + F_{u_i}^{-1}(q) \qquad (1)$$

where $Q_q(PPC_i|x_i)$ denotes the qth conditional quantile function of the potential production of cocaine (PPC) for cristalizaderos i given a set of cristalizaderos specific variables x_i. These specific variables are represented by the

Table 6: Cluster by region of seizure

Region	Cluster 1 %	Cluster 2 %	Cluster 3 %
Meta-Guaviare	42.86	4.03	4.55
Pacific	35.71	37.5	48.48
Putumayo-Caqueta	0	2.02	4.55
Central	7.14	38.31	28.79
Orinoco	7.14	7.66	9.09
Sierra Nevada	7.14	4.03	1.52
Other	0	6.45	3.03
Total	100	100	100
N	14	248	66

Table 7: Cluster by counteracting unit

Counteracting unit	Cluster 1 %	Cluster 2 %	Cluster 3 %
Armada Nacional	28.57	7.26	9.09
DAS	14.29	0.4	1.52
Ejercito	28.57	54.03	50
Policia Nacional	28.57	38.31	39.39
Total	100	100	100
N	14	248	66

amount of both Liquid and Solid Precursors (SP and LP), the number of Ovens (O) and the number of People working in the cristalizaderos (P). We provide other specifications of the model by adding control variables relating to the counteracting unit conducting the seizure (model (2) in Table 9), area where the seizure has been accomplished (model (3) in Table 9) and the month when this has taken place (model (4) in Table 9). This is done in order to control for other factors that may have contributed to explain the productive capacity at the level of the single cristalizaderos. Finally, $F_q^{-1}(u_i)$ represents the inverse of the distribution function of error term u_i.

The quantile regression model has many advantages compared to standard OLS estimation method and it fits particularly well to the dependent variable we want to model, i.e. the potential production of cocaine hydrochloride. In particular, four main advantages are worth stressing: (i) it is robust when the dependent variable departs heavily from normal distribution and in the presence of severe outliers; (ii) a richer characterization of the data is possible given that the impact of covariates on the full distribution, as well as on specific percentiles, of the dependent variable can be investigated; (iii) the assumption of a linear conditional quantile function can be relaxed and departure from linearity can be confidently accepted (Angrist and Chernozhukov, 2006). All of the above

Table 8: Cluster by month of seizure

Month	Cluster 1 %	Cluster 2 %	Cluster 3 %
January	14.29	6.85	12.12
February	28.57	9.68	7.58
March	21.43	6.85	6.06
April	0	7.66	21.21
May	7.14	9.27	9.09
June	0	5.65	6.06
July	7.14	10.08	6.06
August	14.29	5.65	10.61
September	0	11.69	9.09
October	7.14	14.52	10.61
November	0	8.47	1.52
December	0	3.63	
Total	100	100	100
N	14	248	66

mentioned reasons are specifically useful when we come to model the information concerning 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 thanks to the presence of a non negligible number of highly productive cristalizaderos (see the values of skewness and kurtosis in Table 4). Second, for many observations our dependent variable present strong departure from the mean value, thus calling for an evaluation at different points of the conditional distribution, which is possible in a quantile regression framework. Finally, quantile regression allows us to reduce the loss of important information which would have been necessary in more standard settings by dropping outlier observations.

Table 9 shows the results of the quantile regression evaluated at the median value (0.5 quantile), for the four different specifications of the model. Both the extent and significance of the coefficients are robust throughout the different model specifications, showing that the number of people working in a cristalizaderos is significant in explaining the potential production of cocaine hydrochloride, together with the amount of liquid chemical precursors.

Table 10 reports results for the most general model specification throughout the different quantiles of the conditional distribution (column 3 obviously reports the same results as column 4 in Table 9). Other model's specifications report results assigning higher weights to other quantiles of the conditional distribution in this way incorporating the high heterogeneity characterizing the data.

Although the central quantiles of the distribution of cocaine production behave similarly (see columns 2 and 3), the other quantiles present an interesting departure from this result. Indeed, the upper quantiles (columns 4 and 5) reveal

Table 9: Quantile regression (median)

	(1)	(2)	(3)	(4)
Constant	0.82*** (0.194)	-2.89 (1.777)	-1.04 (2.924)	-0.67 (3.035)
Solid precursors	0.14 (0.148)	0.24 (0.150)	0.21 (0.126)	0.21 (0.160)
Liquid precursors	0.0002 (0.0001)	0.0002 (0.0001)	0.0002 (0.0001)	0.0002* (0.0001)
Ovens	0.052 (0.048)	0.047 (0.047)	0.027 (0.066)	0.026 (0.071)
People	0.052** (0.016)	0.053*** (0.013)	0.033* (0.014)	0.037** (0.014)
Counteracting organization dummies				
DAS		5.09 (4.302)	5.67 (3.564)	4.04 (3.519)
Ejercito		3.71* (1.768)	4.00* (1.873)	2.63 (1.879)
Policia Nacional		3.56* (1.721)	4.03* (1.785)	2.65 (1.791)
Geographical dummies				
Pacific			-0.33 (1.974)	0.49 (1.978)
Putumayo-Caqueta			-2.04 (2.470)	-0.21 (2.467)
Central			-2.14 (1.859)	-1.16 (1.780)
Orinoco			-2.95 (1.894)	-2.64 (1.899)
Sierra Nevada			-2.56 (2.527)	-1.94 (2.735)
Other			-2.67 (2.271)	-1.88 (2.098)
Month dummies				
February				-0.07 (0.946)
March				0.21 (1.396)
April				-0.33 (0.938)
May				-0.89 (1.527)
June				0.83 (1.199)
July				0.24 (1.191)
August				0.005 (0.965)
September				0.91 (1.229)
October				0.72 (1.018)
November				-1.37 (1.006)
December				0.65 (1.204)
Obs	166	166	166	166
Pseudo R^2	0.19	0.22	0.28	0.31

Standard errors in parentheses obtained via the bootstrap method
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the important role played by the typology of counteracting organization seizing the cristalizaderos and the amount of ovens with which it is endowed. Even more interestingly, the most productive units (column 5) show peculiarities which, at least partially, confirm the preliminary results obtained with the cluster analysis. In particular, although almost all the counteracting units strongly influence the amount of cocaine potential production, it looks like that DAS is the best performing one. Furthermore, the location of cristalizaderos into the Pacific region appears to have a strong effect on potential production while, among all the month dummies, the only significant one is the month of November which is characterized by an extremely high negative coefficient, thus paring the result obtained in the time–series analysis.

Finally, by means of the proposed model, 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 organizations. In this way we are able to provide an estimate

Table 10: Quantile regression (all the quantiles)

	(1) q 0.1	(2) q 0.25	(3) q 0.5	(4) q 0.75	(5) q 0.9
Constant	0.70 (3.077)	0.70 (2.957)	-0.67 (3.307)	-2.46 (3.580)	-3.75 (4.442)
Solid precursors	0.10 (0.090)	0.06 (0.107)	0.21 (0.161)	0.20 (0.332)	0.31 (0.411)
Liquid precursors	0.0001 (0.00008)	0.0002* (0.00009)	0.0002** (0.0001)	0.0003 (0.0002)	0.00004 (0.0003)
Ovens	0.04 (0.045)	0.03 (0.048)	0.03 (0.076)	0.09 (0.120)	0.34* (0.143)
People	0.02 (0.015)	0.03** (0.014)	0.04* (0.014)	0.03 (0.021)	0.03 (0.025)
Counteracting organization dummies					
DAS	-0.37 (3.610)	3.79 (3.559)	4.04 (3.649)	9.81* (4.200)	15.5** (5.897)
Ejercito	0.11 (1.706)	1.91 (1.774)	2.63 (1.964)	4.12 (2.525)	6.34** (3.255)
Policia Nacional	0.24 (1.715)	2.13 (1.771)	2.65 (1.907)	4.01** (2.446)	6.38** (3.252)
Geographical dummies					
Pacific	-0.34 (2.001)	-1.36 (1.845)	0.49 (2.085)	2.89 (2.327)	6.07** (3.191)
Putumayo-Caqueta	-1.79 (2.415)	-1.20 (2.262)	-0.21 (2.722)	-0.068 (3.650)	-1.34 (4.387)
Central	-0.91 (1.818)	-2.33 (1.703)	-1.16 (1.931)	-0.61 (2.034)	0.06 (2.516)
Orinoco	-1.35 (1.931)	-3.19** (1.846)	-2.64 (2.077)	-1.93 (2.307)	-3.85 (2.760)
Sierra Nevada	-1.88 (2.024)	-2.61 (2.093)	-1.94 (2.751)	-1.71 (4.233)	7.15 (5.073)
Other	-1.23 (2.145)	-2.86 (2.011)	-1.88 (2.300)	-0.65 (2.790)	-0.87 (3.363)
Month dummies					
February	0.22 (0.739)	0.02 (0.827)	-0.07 (0.925)	0.87 (1.695)	3.50 (2.330)
March	-1.09 (1.585)	-0.16 (1.474)	0.21 (1.642)	0.73 (2.130)	-1.07 (2.635)
April	0.25 (0.821)	-0.40 (0.889)	-0.33 (0.950)	-0.42 (1.487)	-1.53 (1.678)
May	0.49 (0.793)	-0.37 (0.914)	-0.89 (1.849)	0.96 (3.379)	3.38 (4.551)
June	0.96 (1.120)	0.57 (0.953)	0.83 (1.191)	0.66 (2.542)	-0.67 (3.737)
July	0.95 (0.834)	0.44 (1.023)	0.24 (1.329)	0.71 (3.639)	-0.78 (5.362)
August	-0.37 (0.808)	-0.15 (0.858)	0.005 (0.985)	0.22 (1.489)	-1.36 (1.649)
September	0.15 (0.753)	-0.58 (0.826)	0.91 (1.216)	1.48 (1.534)	0.32 (1.882)
October	0.39 (0.924)	0.35 (1.044)	0.72 (0.989)	1.27 (1.594)	1.28 (1.820)
November	-1.74 (1.530)	-1.04 (1.244)	-1.37 (0.986)	-1.63 (1.426)	-3.30** (1.717)
December	0.96 (0.820)	0.47 (0.908)	0.65 (1.225)	0.05 (3.433)	0.02 (2.221)
Obs	166	166	166	166	166
Pseudo R^2	0.27	0.27	0.31	0.39	0.53

Standard errors in parentheses obtained via bootstrap method
* $p < 0.05$, ** $p < 0.10$

for all the 328 cristalizaderos, thus also the ones lacking an estimate provided by the seizing authorities. Before presenting the results of the estimation procedure we set forward some remarks concerning the way in which the procedure has been carried out. 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 hydrochloride production for the year 2008 that we propose here is highly conservative in nature. First, as evidenced in the first part of the section, we adopted a quantile regression framework which takes into account the different quantiles of the conditional distribution of the potential production of cocaine which make use of the conditional expected value. We are in this way able to provide a more cautious estimate given the distribution of our data where the mean is always higher than the median (see Table 1). Second, when computing the estimate in order to evaluate the final production of year 2008, we rely on the regression corresponding to the 0.5 quantile. It is worth stressing that adopting the 0.9 quantile would be more correct given the higher Pseudo-R^2 statistic, but it would have inflated the estimate thus yielding an even higher result of total cocaine production. Third, we propose three different scenarios relating to the total amount of cocaine produced in Colombia during the year 2008: (i) one that is 1.96 standard errors lower than the punctual estimate, (ii) 1 standard error lower and (iii) equivalent to the punctual estimate. Fourth, it must be recalled that we take into consideration only those cristalizaderos that have been seized and reported by Colombian authorities.[9] Finally, if more than one cristalizaderos is found to be seized in the same locality, then we assume that any additional cristalizaderos 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 months.

Given these conservative assumptions, the final numbers of our estimation exercise are quite interesting and, we believe, shed new light on the way in which cocaine supply should be conceived by the relevant literature. In the most conservative scenario, that is when the punctual estimate is subtracted 1.96 times the standard error of the estimator, we find a total of 720 metric tons of cocaine hydrochloride produced. It is worth noticing that this number is much higher than the amount of cocaine reported by the United Nations Office on Drugs and Crime (UNODC) report (UNODC, 2009) and the US Government White House Office of Drug Control Policy (ONDCP) report (ONDCP, 2009), which amounts respectively to 430 and 535 metric tons.[10] Furthermore, if only one standard error is subtracted from the punctual estimate then a total estimate of 1500 metric tons is found. Finally, when the punctual estimate is taken as the correct one, then a figure of 2630 metric tons of cocaine hydrochloride is found.

6 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 following the trend. In an attempt to overcome this deficiency in data collection, we put forward a unique dataset covering cocaine related seizures in Colombia for the whole year 2008. By means of this dataset, we offer some preliminary accounts of the main numbers on drug supply and on drug seizures.

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, this effort is aimed at complementing, rather than challenging, the more standard methodologies, in order to provide a more realistic and reliable view of such a complex phenomenon.

[9] It is the case that not all of the productive cristalizaderos are seized by the the counteracting organizations, otherwise Colombia will produce no cocaine at all, while it is still internationally recognized as the major producer of cocaine in the world. Furthermore, it may be the case that the most productive cristalizaderos are actually seized by other counteracting units through international operations taking place just outside the Colombian borders.

[10] The figure of the ONDCP report refers to year 2007, given that no estimate is reported for year 2008.

The dataset seems to be immune from biases due to a sort of strategic use of seizures by the counteracting organizations. Indeed, an overestimate might arise from the desire to inflate results in order to appear more efficient. However, data are stationary, with no structural break suggesting the existence of such a phenomenon[11], which, if interpreted along these lines, seems to confirm that there is a trade–off between seizing too much and too little.

We have indeed shown that the dataset can be of help in offering some novel insights for the sake of counteracting policies. These results are evident from the time–series analysis and from the cluster analysis. Moreover, a model uncovering some of the most prominent features of cocaine production process has been set forward and a preliminary estimate concerning the supply of cocaine in year 2008 for Colombia has been computed. Although the estimate is highly conservative, it is nevertheless much higher than the official numbers provided by the main international organizations.

References

References

Angrist, J., and V. Chernozhukov (2006): "Quantile regression under misspecification, with an application to the US wage structure," *Econometrica*, 74(2), 539–563.

Arkes, J., R. Pacula, S. Paddock, J. Caulkins, and P. Reuter (2008): "Why the DEA STRIDE data are still useful for understanding drug markets," Discussion paper, NBER WP 14224.

Caliński, T., and J. Harabasz (1974): "A dendrite method for cluster analysis," *Communications in Statistics-Simulation and Computation*, 3(1), 1–27.

Caulkins, J., P. Reuter, and L. J. Taylor (2006): "Can supply restrictions lower price? Violence, drug dealing and positional advantage," *Contributions to Economic Analysis & Policy*, 5, 1–18.

Caulkins, J. P., P. Reuter, M. Y. Iguchi, and J. Chiesa (2005): "How Goes the "War on Drugs"? An Assessment of U.S. Drug Problems and Policy," Discussion paper, Rand Corporation Occasional Paper.

Chumacero, R. A. (2008): "Evo, Pablo, Tony, Diego, and Sonny - General Equilibrium Analysis of the Illegal Drugs Market," Working Paper Series 4565, World Bank Policy Research.

Costa Storti, C., and P. De Grauwe (2009): "The cocaine and heroin markets in the era of globalisation and drug reduction policies," *International Journal of Drug Policy*, 20(6), 488–496.

Donati, A. (2009): "La produzione e i sequestri di cocaina. Analisi dei Report ONU dal 1999 al 2007," mimeo.

[11] Apart from the October–December break, which is symmetrical and seems to be due to problems internal to the Army.

ELLIOTT, G., T. ROTHENBERG, AND J. STOCK (1996): "Efficient tests for an autoregressive unit root," *Econometrica*, 64(4), 813–836.

FOWLER, T. B. (1996): "The international narcotics trade: Can it be stopped by interdiction?," *Journal of Policy Modeling*, 18(3), 233–270.

HOROWITZ, J. (2001): "Should the DEA's STRIDE data be used for economics analyses of markets for illegal drugs?," *Journal of the American Statistical Association*, 96, 1254–1271.

IBANEZ, M. (2006): "Who crops coca and why?," mimeo.

KWIATKOWSKI, D., P. PHILLIPS, P. SCHMIDT, AND Y. SHIN (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54(1-3), 159–178.

LEVITT, S. D., AND S. A. VENKATESH (2000): "An Economic Analysis of a Drug-Selling Gang's Finances," *Quarterly Journal of Economics*, 115(3), 755–789.

MEJIA, D. (2008): "The war on illegal drugs producer and consumer countries: a simple analytical framework," Discussion paper, CeSifo Working paper n. 2459.

MEJIA, D., AND C. E. POSADA (2008): "Cocaine production and trafficking: what do we know?," Working Paper Series 4618, World Bank Policy Research.

MORENO-SANCHEZ, R., D. KRAYBILL, AND S. THOMPSON (2003): "An econometric analysis of coca eradication policy in Colombia," *World Development*, 31, 375–383.

NELL, J. (1994): "The dynamics of the drug market," *Challenge*, 37, 13–21.

NG, S., AND P. PERRON (2001): "Lag length selection and the construction of unit root tests with good size and power," *Econometrica*, pp. 1519–1554.

ONDCP (2004): "The Price and Purity of Illicit Drugs: 1981 Through the Second Quarter of 2003," Discussion Paper Publication Number NCJ 207768, Washington DC: Executive Office of the President.

——— (2009): "2009 International Narcotics Control Strategy Report," Discussion paper, 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*, pp. 14–23.

REUTER, P., AND V. GREENFIELD (2001): "Measuring global drug markets. How good are the numbers and why should we care about them?," *World Economics*, 2, 159–173.

RITTER, A. (2006): "Studying illicit drug markets: Disciplinary contributions," *International Journal of Drug Policy*, 17(6), 453–463.

TABARES, E., AND R. ROSALES (2005): *Políticas de control de oferta de Coca: "la zanahoria" y "el Garrote"*. CEDE, Centro de Estudios sobre Desarrollo Económico, Facultad de Economía, Universidad de los Andes.

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.

Thuomi, F. (2003): "The numbers' game. Let's all guess the size of the illegal industry!," Crime and Globalisation Paper, Transnational Institute.

UNODC (2009): *2009 World Drug Report*. United Nations Office on Drugs and Crime.

Van Ours, J., and S. Pudney (2006): "On the economics of illicit drugs," *De Economist*, 154(4), 483–490.

VVAA (2006): "Paco Unders Scrutiny. The cocaine base paste market in Argentina, Uruguay and Brazil," TNI Briefing Series, Debate papers 14, Transnational Institute.

Wilson, L., and A. Stevens (2009): "Understanding Drug Markets and How To Influence Them," Discussion Paper Report n. 14, The Beckley Foundation Drug Policy Programme.