Parallel Foreign Exchange Markets in the 21st Century?
The Venezuelan Case*

José Luis Saboin

INTRODUCTION
The world seems to have abandoned so many things in the last couple of decades: at the end of the 1980s -beginning of the 90s- a lot of things that today are considered as hallows were particularly common. While the second globalization came into full swing during those days and the people started to substitute typewriters, Polaroid cameras and Walkman’s towards more efficient gadgets, lagged economies, in order to achieve macroeconomic stability, were quite opening themselves not only to international trade but towards more market-oriented systems to favor the provision of good and services. In this sense, one hallow that economic science has to its credit (or debit) is the existence of parallel foreign exchange markets.

Just to refresh our memory, parallel foreign exchange markets (PFXM) – which were extremely common in developing countries half a century ago- are those in which a market-determined exchange rate coexists with one or more pegged exchange rates.

As good hallows, PFXM not only possess an extensive theoretical and empirical literature on the macroeconomic implications of a parallel foreign exchange rate, but they also possess real life experiences. Whereas the number of –prop-
erly identified– PFXM was 41 by 1983, there were 14 by 2003 and, by 2013, just 4. Despite there are a great variety of experiences one common characteristic is that the transactional premium in such markets tends to be quite high2.

Empirical evidence reveals that PFXM are adopted following two patterns. First, as a response to a balance of payments crisis, its objective is to isolate the typical overshooting of the foreign exchange rate effect on domestic prices while maintaining limited control over international reserves. The second wave shows that a (black) PFXM develops gradually in response of widespread restrictions in the official (appreciated) exchange rate market. The premium generally rises in the latter when a balance of payments crisis occurs.

While cross national studies are important to evaluate different hypothesis about the causes and consequences of an economic phenomenon, there has been a growing awareness about the limitations of cross country economic data to help us understand in depth the particularity of some of cases at the individual level. To assess this issue, Rodrik (2005) and others tried to form a broader picture using a from-bottom-to-top analysis. In this sense, while my interest is to understand why this phenomenon raised again the 21st century at a broader level, the purpose of this paper –without losing the generality of the issue– is to analyze the phenomenon at the particular level, starting with the case of Venezuela, leaving as a medium term goal the cross country analysis and final conclusions.

Venezuela, by its peculiarity of oil exporting country, has a long tradition with foreign exchange controls. Its experience goes back to the 1960s when an oil-related disturbance affected the balance of payments and the government, successfully, applied capital controls to smooth the oil shock on domestic prices. Later, from 1983 to 1989, when the Latin American debt crisis hit and the country faced balance of payment problems the government enacted capital controls again but the experience was not satisfactory at all. Then, after a successful implementation of a floating exchange rate regime and capital mobility in the early 1990s, a financial crisis that affected the whole domestic payment system was the primarily cause to enact capital controls from 1994 to 1996. Finally, at the beginning of 2003, a political

1 See IMF book on exchange rate arrangements, several years.
2 Kiguel and O’Connel (1994).
conflict that triggered in the state-owned oil company (Pdvsa) to decrease its production levels to a point in which foreign currency income and international reserves reached critical levels, was the excuse of the authorities to impose a capital control that continues to this day.

One issue that particularly calls the attention, beyond the fact that 21st century Venezuelan policymakers repeated the same set of policies after the bad experiences of the 80s and 90s, is that the last capital control is the longest lived (11 years so far), the one that reported the highest premium (1,381%), the one with the largest amount in capital flight (US$158 billion) and the one with an inflation rate that has averaged 25% over the period. However, a much more complicated issue arise when the question of how the authorities allowed the premium to reach those levels is formulated.

While the results of this research points out that the answer is an interesting topic for a broader research including political economy analysis, I argue that the long span of time of this last exchange control has to do with two important patterns that works in tandem. One has to do with Venezuela’s oil tail, particularly the part associated to the control of the oil income by the State and the other has to do with the political situation of the country where an excess of electoral events unleashed expansionary and unsustainable economic policies.

The paper continues as follows. In section 2, I will discuss the background and characteristics of a parallel exchange rate system and its Venezuelan form. Focusing on the adoption of such system in the country I have found that it has been used in both ways: (i) to insulate prices and the capital account and (ii) to maintain an overvalued exchange rate. In this matter, I argue that the government has failed in the first one but succeeded (if that it so) in the second. In section 3, I discuss the functioning of a parallel foreign exchange market using a simple analytical framework of the stock and flow model developed by Dornbusch et. al. in 1983 where the parallel exchange rate (and the premium) is jointly determined by current and asset market conditions. Then I use an econometric specification to test if the main theoretical arguments apply in the Venezuelan case and compare it with other empirical evidence. Section 4 derives in some stylized facts about

---

the last Venezuelan exchange control on the balance of payments and domestic prices. Section 5 discusses the unification issue. Section 6 delivers the fiscal paradox of parallel exchange markets in oil exporting countries and Section 7 concludes with the lessons learned.

BACKGROUND

What type of PFXM are we talking about?
PFXM have been roughly classified according to two characteristics: coverage and legality. Coverage is referred to the way in which parallel markets are used for. Generally, this system works with two types of transactions: one in which the free rate uniquely applies to financial account transactions and other in which, besides financial transactions, a broad range of current account transactions are allowed at the free rate.

Regarding lawfulness, the literature classifies parallel markets into two subcategories, legal and illegal. Given the high costs of enforcement, governments typically tolerate a substantial amount of illegal parallel market activity, but while it is usual to observe attempts to suppress parallel markets, succeeding in doing it however is quite unusual or, at least, short-lived. In this sense, we can distinguish between systems that are legal or highly tolerated -but illegal- and those in which a threat of enforcement is present during a period of time. Based on this classification, historical evidence shows that illegal parallel markets are short-lived while legal parallel markets are long-lived.

Current and past experiences in Venezuela reveal that the country has passed by all the classifications. The first exchange control (1960-1963), can be classified as a legal parallel market functioning only for capital account transactions. The second exchange control (1983-1989), known as RECAIDI by the Spanish acronym Regimen de Cambios Diferenciales (Differential Exchange Rates Regime), could be introduced within the category of legal systems supported by severe foreign exchange rationing for imports. The third exchange control (1994-1996), known as OTAC because of the acronym Oficina de

---

4 In practice, coverage and legality are more matters of degree than discrete categories.
5 Kiguel and O'Connel (1994).
Administracion Cambiaria (Exchange Office Administration), could be also within the category of legal systems, but this time without restrictions on imports.

The fourth and actual exchange control (2003- ), known as Cadivi because of the Spanish acronym Comision de Administracion de Divisas (Foreign Currency Administration Commission) has passed through all the classifications. From 2003 to 2008, it was a totally legal system used for capital account transactions and a marginal portion of current account transactions, principally imports of non-essential goods and services. From 2009 until May 2010, after the global financial crisis of 2008, the proportion of allowed current account transactions increased slightly and the system continued to be completely legal but with a premium increase ranging from 30% to 45% with respect to the 2003-2008 period. Between May 2010 to August 2012 however, just after a timid 21% increase in the premium, parallel market transactions were severely penalized and around 10% of imports were removed out of it. By February 2014, after a premium increase of 508% with respect to August 2012 the parallel market was declared licit again amid a huge increase on the scarcity index reported by the Central Bank of Venezuela (BCV)⁶.

<table>
<thead>
<tr>
<th>Period</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 1964 - Jan 1983</td>
<td>Fixed ER and Free Capital Mobility</td>
</tr>
<tr>
<td>Feb 1983 - Jan 1989</td>
<td>Multiple ER and Capital Controls (Recadi)</td>
</tr>
<tr>
<td>Feb 1989 - Sep 1992</td>
<td>Floating ER and Free Capital Mobility</td>
</tr>
<tr>
<td>Jul 1994 - Jul 1996</td>
<td>Multiple ER and Capital Controls (OTAC)</td>
</tr>
<tr>
<td>Aug 1996 - Jan 2002</td>
<td>Crawling Peg</td>
</tr>
<tr>
<td>Feb 2002 - Jan 2003</td>
<td>Managed Float</td>
</tr>
<tr>
<td>Feb 2003 - Today</td>
<td>Fixed ER and Capital Controls (Cadivi)</td>
</tr>
</tbody>
</table>

**Adoption of the parallel market**

In essence, there are two broad patterns in which a parallel foreign exchange market arise and become important in an economy. In the first pattern, the policymakers divide the foreign exchange market in order to phase in a

---

⁶ Scarcity Index reached to 30% in December 2013 while stress levels declared by the BCV are around 14%.
devaluation when capital outflows prompt a balance of payments crisis. In the second pattern, the parallel market emerges gradually in response to efforts to maintain an overvalued exchange rate; in this case, policymakers become forced to restrain access to the official market for both financial and current account transactions and, eventually, controls are tightened and the illegal market begins to gain macroeconomic importance.

In a long tradition with capital controls, Venezuelan governments have had the incentives to adopt the two main patterns in different occasions. However, the last adoption contains elements related not only to the first but to the second pattern as well. In this sense, it seems there could be other motivations for the use and maintenance of the last system for such a long period which are going to be discussed more forward.

Pattern 1: Inflation control and stemming capital outflows

In this scenario, dual exchange rate systems are generally adopted when a balance of payment crisis required a devaluation of the currency with high implications on inflation. The theoretical justification behind this argument is straightforward. In the case of a fixed exchange rate regime, a dual system insulates the international reserves from capital outflows, since it leads to a depreciation of the parallel rate instead of a loss of reserves. In floating exchange rate regimes, since current account transactions are done at the official exchange rate (OER) a dual system helps to limit the impact of capital outflows on domestic prices.

In Venezuela, the use of dual rates was modeled in the 1980s taking into account the successful experience of 1960s. In the experience of the 1960s, the authorities had instituted a dual system as a way of phasing in a devaluation of 35% over a period of three years. Based on that experience, in 1983, Recadi was initially adopted as a temporary measure to deal with the lessened oil revenues and massive capital flight. The target was similar to the one of the 1960s: the authorities envisioned a three-year transition to a unified fixed exchange rate regime, with a cumulative devaluation of 40%.

7 Fears of devaluation by the government and rent-seeking probably played a role in this regard.

8 However, oil prices dropped again in 1986 and the parallel market continue functioning until 1989 when it was substituted by a Floating Exchange Rate Regime amid the market reform process known as “Gran Viraje” (Great Turn) by its pretensions to change the State-led development model that was installed in the country since 1940s.
In the current parallel market, it was not the oil price but the drop in international reserves which, after having dropped by 45% between May 2001 and January 2003\(^9\) became the main argument for the implementation of the capital control. By March of that year, Venezuelan policymakers established a capital control. In that juncture, it could seemed necessary the implementation of such measure, however, by November of 2003 international reserves have recovered to May 2001 levels, the equivalent to an increase from three to six months in imports coverage by that time and the political situation cooled down. Therefore, from the macroeconomic point of view it could be perfectly argued that the timing was ideal for the elimination of the control. Notwithstanding, from the political economy point of view, the situation was quite non-ideal, in my view.

At the beginning of 2004, a referendum was called to ask for the resignation of the president. In that moment, the government, in the middle of an oil price boom, maintained the capital control –keeping the exchange rate fixed– and started to increase public spending with social purposes. The expansionary benefits of keeping the exchange rate fixed while increasing public expenditure yielded high political dividends to the government and it won the referendum.

The continuation of this policy during a successive period of elections\(^10\) takes us to the second case for the adoption (though this case maintaining) of a parallel market.

**Case 2: Parallel market to maintain an overvalued ER**

When parallel markets arise, the typical pattern is a gradual worsening in the balance of payments as a result of expansive monetary and fiscal policies that rise inflation and lead to an overvaluation of the OER.

The authorities might fail or might be unwilling to correct those imbalances through a tightening of macroeconomic policies or devaluations of the OER. The economic argument behind this “fear of devaluation” is that

---

\(^9\) During the same period oil exports declined by 2.1 million barrels per day (mb/d), from 2.9 to 0.8 mb/d.

\(^10\) During the last 10 years Venezuela has had 9 elections per year. 4 of them where out-of-calendar.
unification would likely push the exchange rate to overshoot its equilibrium level, producing an important fall in real wages, credit and output.

That being the case, restrictions are gradually tightened in an attempt to resolve the fundamental inconsistencies between the exchange rate and the pressure of expansionary macroeconomic policies. But expectations play a key role here: a possible maxi-devaluation of the OER, or a further tightening of foreign exchange controls, adds to the excess demand for foreign exchange in the short run by encouraging inventory accumulation of importers and portfolio swings from local to foreign currency by investors. Thus, a growing illegal (black) market typically reflects both: a systematic bias against devaluation and a profitable investment strategy by private (or public) investors.

But why a government will be unwilling to correct such macroeconomic disequilibria and keeping an overvalued exchange rate for such a long time? A main argument to answer the question relies on the economic and political advantages of an expansionary fiscal policy amid a fixed exchange rate regime. A secondary argument relies on the increasing purchasing power of the domestic currency in the international markets during such periods and its impact on consumption. Of course, both arguments by themselves do not count as sufficient explanations to answer the question, other factors, such as using the ownership of the oil rent by the State to control and direct the private sector activity should play a key role in assess this question.

From a political economy point of view, the first argument totally makes sense if one states that the incumbent government, near an election episode, would be benefited from an appreciated exchange rate to enhance the multiplier effect of expenditure on aggregate income. This postulate makes more sense in the case of Venezuela since the incumbent government has faced an environment of permanent elections.

The second argument lies behind the latter. Given Venezuela’s oil tail (i.e. the non-oil sector of the economy does not fulfill the requirements of sudden and high increases in demand) a sequence of short-run fiscal stimuli and their consequent impact on consumption (which has to be fulfilled by more and more imports) requires keeping the exchange rate fixed in order to avoid the inflationary pressures of the very short term. Clearly, this practice exacerbates the natural-resource-course condition when subsidized
imports overshoot the non-oil domestic production and adjustments in non-tradable sector prices boost inflation in the long run.

While both practices fall short from a macroeconomic stance in the future since they create several boom busts cycles and (hence) political instability, the point that I want to make clear here is that the implicit political situation plays a key role in explaining the maintenance of an overvalued exchange rate for such a long period time. A similar explanation also applies in the maintenance of the control on capital mobility: during the period of study, capital outflows have accumulated a total of US$158 billion while inflation averaged 25% year on year.

**DETERMINANTS OF THE PREMIUM**

**Theory**
The literature review done by Agénor (1990a, 1992) and Lizondo (1990) is going to be used as Hausmann (1990) and Kiguel and O’Connell (1994) did to adapt an analytical framework based on the stock/flow model of the parallel exchange rate literature\(^\text{11}\) and explain the Venezuelan case.

A simplified version of this model tells that the parallel exchange rate premium is jointly determined by asset market conditions and “parallel” current account conditions. Such asset market conditions play a central role in the short run, up to the extent that portfolio decisions and expectations predict that the parallel exchange rate will depreciate (and the premium increases) in anticipation of a devaluation of the OER and an increase in the money supply. The parallel current account on the other hand, plays a central role in the horizon, i.e. increasing restrictions in official exchange rate imports raises the premium in the long run as imports are diverted to the parallel foreign exchange market.

Following this approach, the model works as depicted in the following figure:

In the short run, the parallel exchange rate is determined by portfolio equilibrium. Since financial transactions take place at the parallel rate, the main determinant of the domestic currency return of foreign assets is the expected rate of depreciation in the parallel market.

Letting B (as “Black”) be the parallel rate, this expected rate of depreciation is given by:

\[ \frac{B_{t+1}}{B_t} - 1 = \frac{E_{t+1}}{E_t} \cdot \frac{Z_{t+1}}{Z_t} - 1, \]

where \( E \) and \( Z \) are the OER and parallel premium respectively, \( Z = B/E \) is one plus the parallel premium and \( t+1 \) are the expected values. The demand for dollars is depicted as the downward slopping function \( DD \) on the right hand side of the figure.

Given the current Premium, \( DD \) increases (shifts to the right) with:

- The expected rate of official depreciation: \( \frac{E_{t+1}}{E_t} - 1 \)
- The expected future Premium: \( Z_{t+1} \)
- A raise in the foreign currency value of domestic assets \( (M/E) \), as wealth holders seek for portfolio rebalancing.
Also, given the premium, \( DD \) decreases (shifts to the left) with a rise in the interest rate on domestic assets, \( i \). The net stock of foreign assets held privately is \( F \) and portfolio equilibrium prevails when this stock is willingly held, this is depicted at point 0, where \( DD \) and \( F \) intersect.

Therefore, in the short run, the premium is:

- An increasing function of the expected future premium, \( Z_{t+1} \)
- An increasing function of the real stock of domestic financial assets, \( M/E \).
- An increasing function of the official interest parity differential

\[
D_t = i^* + \omega \frac{E_{t+1} - E_t}{E_t} - i_t
\]

- A decreasing function of \( F \).

In the long run, the influence on the parallel premium comes from the parallel current account balance (the left hand side apparatus in the figure above). The flow supply of dollars, \( S \), comes from legal exports, exports underinvoicing/smuggling and import overinvoicing. The flow demand of dollars, \( D \), comes from legally assigned imports and import smuggling. A rise in the premium stimulates the flow supply and discourages the flow demand, resulting in an increased parallel current account. The requirement that the parallel current account eventually has to be balanced implies a long run relationship between the premium and other determinants of the parallel current account, such as \( F \), if interest income goes through the parallel market.

Thus, a real appreciation worsens the parallel current account by reducing aggregate exports and increasing the incentive for import smuggling into the domestic market; this drives up the premium in the long run. Similarly, a rise in import tariffs (or a fall in export taxes) worsens the parallel current account by encouraging smuggling and underinvoicing of imports, thus the parallel premium rises.

Since net capital flows at the OER are typically prohibited, the primary source of changes in the stock of foreign assets is the parallel current account. This implies that the flow determinants that apply for the long run can also influence in the short run. Equally, since domestic asset stocks
changes are assumed to be primarily financed through monetization, fiscal variables also play an important role in both, the short and the long run.

Figure 1 shows the long run adjustment of an increase in the domestic money stock measured in dollars under the assumption that E remains fixed. Starting at point 0 a monetary expansion shifts DD to the right and increases the premium to Z₁ in the short run. Under static expectations regarding the parallel rate, the premium subsequently falls as the parallel current account surplus leads to an increase in F. Under rational expectations, market participants foresee the appreciation that occurs in the parallel market during the adjustment path; this reduces the demand for dollars, and adjustment to point 2 takes place along a downward sloping path below DD₁².

The main results from this model come in two fronts. In the long run, adjustments are ruled by the evolution of asset stocks and expectations of depreciation over time. Since the primary source of F is the parallel trade balance, this implies that the determinants that influence the premium in the short run can also influence the premium in the long run.

Monetary financing and the fiscal balance also have a role in both the short and the long run. An increase in real domestic money or in the interest parity differential in favor of F, rises the premium in the short run. Appreciation of the real exchange rate or import restrictions raises premium in the short run and in the long run. The nominal exchange rate affects the premium only indirectly, through real money balances or the real exchange rate, nominal devaluations therefore have no effect on the premium if they are fully offset by money growth and domestic inflation.

**Empirical estimation**

**Hypotheses**
I this section, an empirical specification of the stock/flow model was estimated following the approach of Ghei and Kiguel (1992). In particular, I examine the relationship between the premium, the money stock, the stock

---

12 Following Kiguel and O’Connel (1994) I assume that interest income channeled through the parallel market is negligible so that the monetary expansion has no long-run effect on income. If interest income is important, the accumulation of foreign assets would shift S to the right over time, and the premium would fall in the long run.
of international reserves, and expectations of devaluation of the official exchange rate. Consequent with the discussion above, the following hypotheses were stated. (i) There should be a positive relationship between the parallel exchange rate and the money stock, as postulated by simple monetarist models. Since the relevant definition of money for this purpose is an empirical question, in this case I will use M2. (ii) Large changes in the money supply which are not compensated proportional changes in the official exchange rate will lead to an increase in the premium. (iii) Falling stocks of international reserves could signal the approach of a balance of payments crisis which would cause the parallel rate to depreciate and the premium to rise. (iv) Expectations of devaluation of the OER in the immediate future should lead to depreciation of the parallel rate and an increase in the premium. In some cases, devaluations can be predicted by looking at the evolution of the real official exchange rate: a strong real appreciation could indicate that a devaluation is coming. Thus, a deviation of the real exchange rate from its equilibrium value would then influence the level of the premium.

**Empirical specification**

Given that such macroeconomic variables are likely to evolve along similar paths, a test for cointegration was done. The premium could be said to be cointegrated in the sense of Engle Granger (1987). For the following expression, where $b_t$ is the parallel premium and F is the vector of variables,

$$\ln b_t = \beta' F$$

(1)

if both $b_t$ and F are first difference stationary, then (1) is integrated of order zero and hence, the left hand side describes a cointegration relationship where $\beta$ is the cointegrating vector. Since the residual (estimated error) is integrated of order zero we obtain that there is a stable equilibrium relationship between the premium and the fundamentals. Dickey-Fuller tests have passed for all the series to be first difference stationary. I proceeded to estimate a cointegrating relationship as follows:

$$\ln b_t = \beta_0 + \beta_1 \ln m_t + \beta_2 \ln R_t + \beta_3 \ln \Delta \delta_t + \beta_4 \ln b_{t-1} + \epsilon_t$$

Where $m_t$ is broad money divided by the official exchange rate, $R_t$ is the stock of non-gold international reserves, $\Delta \delta_t$ is the deviation of the real exchange rate from its equilibrium value and $\epsilon_t$ is an i.i.d error term.
The expected signs of the coefficients are:

\[ \beta_1 > 0, \beta_2 < 0, \beta_3 > 0, \beta_4 = ? \]

The approach used for calculating the movements of the OER from equilibrium went as follows. The real effective exchange rate was calculated using the nominal OER, the consumer price indexes of Venezuela’s main trading partners (which account for around the 90% of trade) and Venezuela’s consumption price index. Of course, the equilibrium real exchange rate depends on a number of other variables such as fiscal and monetary policy, the terms of trade, etc. Unfortunately, these variables are not available with the frequency used for estimation. Alternatively, the equilibrium real exchange rate was calculated as a twelve period (month) moving average. The assumption here is simple: the actual real exchange rate observed during a span of time (usually three years) roughly corresponds to the equilibrium one. Thus, the difference between the (calculated) equilibrium real exchange rate and the actual real exchange rate was used to proxy for the extent of disequilibrium of the real exchange rate at any point in time.

\[ \Delta \text{ln} \hat{e}_t = \text{ln} \hat{e}_t - \text{ln} e_t \]

The error-correction equation provides a specification for the short run dynamics of the behavior of the parallel premium and has the following representation:

\[ \Delta q_{t+1} = -B(\hat{e})_t + u_t \]

Where \( \hat{e} \) is the residual from the estimation equation and is a disturbance term.

The estimation period was from 2003 to 2014 using monthly data. All data are from the IFS, Venezuelan Central Bank and the Finance Ministry except the parallel exchange data that was obtained from domestic private think tank Ecoanalitica based on secondary sources. The results for the cointegration and error-correction models are presented below.

**Empirical results**

The following table presents a summary of the econometric specification. The last two columns show cointegration and error-correction results, the
first two, on the other hand, show the same specification but with the purposes of testing the cointegrating relationship between the real exchange rate and the parallel market rate.

As expected, column 4 shows that money supply, measured at the OER, has the expected sign and is significant at the 1% level. The coefficient obtained for international reserves has the expected sign but it is not statistically significant at the conventional levels. The coefficient obtained for the term representing deviations from equilibrium of the real exchange rate has the expected sign and is statistically significant at the 1%. All ADF tests strongly supports the hypothesis of cointegration.

Short run dynamics, explained by the error-correction equation of column 4, suggests that disequilibrium in the market is primarily influenced by the money stock and depreciation expectations (both significant and with the expected sign). The residual term is negative and statistically significant, providing support for the hypothesis of stable dynamics.

<table>
<thead>
<tr>
<th>Table 2. Regression results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Estimation Method</td>
</tr>
<tr>
<td>Premium</td>
</tr>
<tr>
<td>Parallel Exchange Rate</td>
</tr>
<tr>
<td>Money Supply</td>
</tr>
<tr>
<td>International Reserves</td>
</tr>
<tr>
<td>Depreciation Expect.</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Residual</td>
</tr>
</tbody>
</table>

The results are consistent with theory as the relevance of money balances (measured in $ at the OER) rising the premium. Also, money growth in excess of the rate of official depreciation rises the premium. Given the link
between the budget deficit and money growth, the persistency of a high premium shows the inconsistency between the public sector budget deficit and exchange policy in Venezuela, especially, during the last four years.

Since theory predicts that interest rate parity differentials increase the premium and Venezuela has a long tradition of negative real interest rates, interest parity has become dominated by exchange rate expectations. In this regard, the sign and significance of the coefficient that measures depreciation expectations is highly significant.

Regarding devaluation, the following graph shows the premium and devaluation during the studied period. It reveals the cycle of official devaluations. While in previous periods the effects on the premium of a devaluation of the OER are negative but very transitory, the evidence of the last two years suggests that in the absence of supporting macroeconomic policy management over the fundamentals, the effects on the premium of a devaluation of the OER are muted.

<table>
<thead>
<tr>
<th>Figure 2. Devaluation and premium</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Graph showing premium and devaluation over time." /></td>
</tr>
</tbody>
</table>

Sources: Central Bank, Ecoanalítica and own calculations.

The following graph shows how real appreciation of the exchange rate (measured as the overvaluation of the OER) increases the premium in both the short and the long run.
The remaining specification was estimated with the intention of testing cointegration among the real exchange rate and the parallel rate. This will allow us to see not only if the law of one price holds when a parallel market arises, but also to see if the parallel market moves along with the fundamentals of the economy, namely, TOT, tradable sector productivity, domestic supply and demand, etc.

The highly significant coefficients and positive signs of the parallel exchange rates expressed in the regressions within the first and second columns of Table 2 shows that there is a cointegration relationship in both the short and the long run between the real and the parallel exchange rate.
Figure 4. **Real and parallel exchange rates**

[Diagram showing real equilibrium exchange rate and parallel market rate over time, with VEB/US$ on the y-axis and dates from Feb-03 to Feb-14 on the x-axis.]

**Sources:** Central Bank and own calculations.

**STYLIZED FACTS**

This section proceeds with the analysis of the empirical results but contrasting them with other empirical findings in order to derive some stylized facts regarding the impact of a parallel foreign exchange market on the balance of payments and inflation.

**There is limited insulation of the BOP**

In theory, dual and multiple exchange rate systems insulate international reserves from capital flows and portfolio shifts while allowing the authorities to maintain full convertibility for current account transactions. In practice, this is not possible because the size of the premium affects the official trade flows stimulating leakages in such transactions making the authorities to fully separate capital and current account transactions. Therefore, one could argue that insulation is a declining function of both, the size of the parallel premium and the length of time that the system remains in place.

In this sense, previous empirical evidence contributes explaining the PFXM impact on the BOP in its Venezuelan form. For instance, Kamin (1991) found that agents resorted to overinvoicing of imports and underinvoicing of exports in order to gain access to the OER. Marion (1992) found that in French and Italian PFXM episodes agents used trade credit to move capital in or out...
of the countries. Finally, Kaminsky (1991) shows the capital flight in Mexico was higher after the dual system because Mexican authorities intervened in the parallel market to avoid further depreciation of the parallel rate. This experiences show that the imposition of stricter foreign exchange controls or the “too-little, too-late policy” only worsens the picture by increasing the incentives of agents to obtain foreign currency at the cheapest rate available in the economy leaving the central bank in a vulnerable position; while devaluation, ease, tolerance or elimination of controls does the opposite.

For instance, during the illegalization phase of 2010-2013 which was casually the period in which the premium raised the most BCV’s international reserves levels dropped to their historical minima. Moreover, if one compares international reserves levels between periods of legality one can observe substantial differences (see Figure 5). Thus, the illegalization period raises an important fact: although the exchange rate was nominally pegged and import controls were tighter, both international reserves and imports reached undesired levels affecting the BOP.

**Figure 5. Non-gold international reserves**

![Graph showing non-gold international reserves over time.](image)

*Sources: Central Bank and own calculations.*

One possible counter-factual to the above could be the argument that, as consequence of the Venezuelan natural resource course exacerbation during the Manna from heaven period (a period in which the nominal ex-
change was pegged the whole time), the level of absorption of tradable goods (but also the parallel current account) of the economy increased significantly. Between 2003 and 2008 imports of goods increased by a cumulated factor of 3.5.

So, taking into account that a non-negligible part of imports was increasingly attended by the parallel market, and that the parallel market itself worked as an exhaust valve for other current account transactions, its illegalization became a heavy burden to the Central Bank. Or in more simple words, any agent within the Venezuelan economy was after the most precious and needed good, which was not the pursuit of happiness, but the cheap dollar, the imports dollar, the Cadivi dollar.

Another explanation could be related to the incentive that the premium generates between bureaucrats in the foreign exchange regulatory institutions and private and public economic agents to arbitrate the system. The banning of the parallel market –the rise in the premium– increased this incentive to the point that some agents and bureaucrats falsified business in order to obtain the cheap (official) dollars. Moreover, in an environment of low quality of governance such as Venezuela14, this practice became commonplace, increasing the aggregate demand for dollars and distorting the allocation of resources. The misallocation of resources in turn resulted in shortages of goods and services.

13 This is referred to the period during 2004-2008 years when the real oil price for the Venezuelan basket six folded.
14 According to WB Governance indicators 2012 database Venezuela occupies the percentile 7 from zero (lowest) to 100.
Given the high –economic and political- detection costs of briefcase companies, an expeditious solution to this issue which is not associated to complete unification seems quite difficult to be successful.

**There is temporary insulation of domestic prices**

History indicates that parallel markets can at best provide temporary insulation of domestic prices, specially, when the pressures for devaluation are generated by short term capital outflows. At times of balance of payments problems the exchange rate tends to be extremely volatile usually overshooting its long run equilibrium level, in such cases, a dual exchange rate system can limit the inflationary impact of depreciation by allowing the financial rate to absorb a major share of the balance of payments pressure.

In the case of Venezuela the study of this impact requires deeper analysis because price controls in some basic goods (food and gasoline) were imposed in tandem with capital controls. However, at the beginning of the controls we can clearly observe how the parallel exchange rate depreciated drastically, overshooting its eventual long run equilibrium level and insulating prices (see Figure 7). However, as well as in most of PFXM episodes, the real parallel exchange rate starts to appreciate around three months and stabilizes at nine months after the crisis.
Additionally, a PFXM’s scope is much more limited when the pressures for devaluation have to deal with problems resulting from generalized excess demand. Evidence shows that in countries that fail to control monetary growth, the parallel market depreciates and domestic inflation rises even if the government keeps the exchange rate fixed\textsuperscript{15}. As empirical results have shown, the Venezuelan experience is quite relevant in this matter: nowadays despite no significant balance of payment episode, public sector deficit has increased to the point that central bank has started to lend money to public corporations. As a result, inflation has mounted since the last quarter of 2012.

\textsuperscript{15} See Kaminsky (1991) and Ghei and Kiguel (1992).
Figure 8. Non-financial public sector result

Sources: Central Bank and own calculations.

Figure 9. Monetary financing

Sources: Central Bank and own calculations.
TOWARDS UNIFICATION

Evidence
Two interpretations of the term unification are important in practice. The first refers to the adoption of a single exchange rate for all external transactions, with full convertibility, this is known as “full unification”. The second refers to the adoption of a single rate for all current transactions, while maintaining convertibility restriction and therefore a parallel market for portfolio and other financial operations. This is known as gradual unification.

Full unification
Paradoxically, full unification is often adopted at times of macroeconomic crisis. Evidence shows that this is a particular phenomenon in Latin America with cases such as Mexico, Argentina and Venezuela doing it in several occasions. As a matter of fact, after six years of operating a multiple system, Venezuelan authorities, fully unified the foreign exchange market in February 1989 by floating the exchange rate amid balance of payments problems16. However, that unification was part of a successful macroeconomic stabilization program that included a reduction in the fiscal deficit, trade liberalization and the introduction of market mechanisms. Moreover, in Venezuela, unification (as intrinsic devaluation) exerts an important positive effect on the government’s budget.

Gradual unification
This a phenomenon of highly distorted economies. Evidence shows different patterns with a common element: the compromise to sustain the unification. In the case of Turkey, for example, the process lasted nearly a decade. Its first phase was concentrated in the unification of current account transactions, while keeping the premium small. Secondly, the long term goal was to achieve unification. This was obtained by the adoption of a more flexible exchange rate management so as to maintain a realistic RER, liberalization of imports, and relaxation of controls on the capital account until its full liberalization. Also, the process was accompanied by more fiscal and monetary discipline.

Two African countries examples could shed some light on the Venezuelan situation: Ghana and Tanzania in 1980s. In the case of Ghana unification pro-

16 By that time, the premium was close to 200%.
ceeded gradually as part of a comprehensive reform process that began in 1983, which included monetary and fiscal restraints, increases in producer prices, relaxation on import controls and more flexible management of the exchange rate. At the begging the foreign exchange market was transformed into one that comprised two legal markets with floating exchange rates and a negligible spread between them, such system coexisted with a reduced illegal market.

Results were satisfactory: black market premium declined from above 2000% in 1983 to 15% in 1988, when the second legal market became operational and practically absorbed the black market. The spread between the two legal markets declined gradually and reached around 5% in April 1990 remaining low until financial account liberalization—and therefore full unification—in 1994.\footnote{Ghana exchange rate structure was changed back to dual on July 2012.}

In the case of Tanzania, when partial unification started the premium declined from over 700% in 1986 reaching 50% in 1990. This happened in three steps. In 1984, the authorities devalued and introduced an “own-funds” scheme that allowed holders of (illegal) foreign exchange to obtain import licenses freely. By 1986 this scheme was financing a third of total imports. In that year, the authorities devalued again and adopted a crawling peg system in tandem with a major macroeconomic reform package. In 1992, foreign exchange bureaus were created and foreign currency deposit accounts in domestic banks were introduced. In mid-1993, full unification was achieved. By 1994 an interbank foreign exchange market replaced the bureau system and requirements on export proceeds and other export restrictions were gradually removed in the coming years.

Failures and successes
History shows that countries with large budget deficits financed by money creation that kept used the OER as anchor for inflation are the ones that failed when trying for unification. History also reveals that success relies in the acceptance of the need for consistency between the unified exchange rate and monetary and fiscal policy.

In practice, exchange rate policy has to accommodate the underlying inflationary pressures generated by the fiscal deficit. Fiscal discipline, tightening
in domestic credit and inflation management are the main policies that the countries that committed to unification applied.

Evidence also shows that economies with extensive price controls, barriers to trade, and thin financial markets required a gradual approach. In Africa, for example, structural reforms aimed at enlarging the role of market mechanisms in determining resource allocation, this was determinant. However, in my opinion, it was the commitment by the authorities towards unification and the further adoption of a sustainable exchange rate regime what in the end made all possible.

THE FISCAL “TAIL”

Given the country’s oil specialization –around 54% of the government revenues and 96% of exports are related to oil income– it is straightforward to see the relationship between fiscal policy and exchange rate policy decisions. In this section, using rough estimates, I will estimate the fiscal (and quasi-fiscal) consequences of parallel exchange rates in Venezuela.

The following equation summarizes the fiscal impact of parallel exchange rates:

\[(D^*-D) + [(E^*-E).(x-m)]^{18}\]

The star (*) over each variable indicates the value it would take in a hypothetical unification. Thus, \(E^*\) is the exchange rate that would be consistent with equilibrium in the BOP, or the “shadow exchange rate”. \(D\) denotes non-financial public sector deficit. \(X\) and \(M\) denotes CB purchases and sales of foreign exchange. The first term represents the effect of the multiple system on the non-financial financial public sector’s budget. The second term shows the implicit gains or losses incurred by the central bank purchases and sales of foreign currency.

The components results, however, do not necessarily work in the same direction. In the case of Central Banks profits it will depend if it is a net buyer or a net seller to the private sector. Since the OER is usually appreciated

---

18 Kiguel and O’Connell (1994).
against the parallel, net sales to the private sector generate losses while net purchases generates profits. Non-financial public sector budget depends to the extent on how many components of the budget are indexed at the OER.

Applying this methodology to the Venezuelan case, for simplicity -but without loss of generality- I will assume that the shadow exchange rate is equal to the parallel exchange rate. Since the Central Bank is a net seller of foreign currency and despite the buying and selling rates are different, the calculated losses of the Central Bank reached 13 GDP points on average during the period. It is acknowledged that the size of this estimates may be biased upwards since the parallel exchange rate most likely overestimate the true value of the foreign exchange. However, the estimated effects remain significant after correcting for possible bias.

In the case of the non-financial public sector deficit, Venezuela’s public sector is a net producer of foreign currency, as revenues from oil exports far exceed the cost of servicing the public external debt. However, oil-related contributions, custom duties, purchases of foreign goods, sales taxes on imports among other minor items, are fully indexed to the OER, while other items, such as other tax revenues and public wages are not. In this sense, multiple rates could have both, positive and negative effects on the domestic currency budget.

It is also important to highlight the impact this also have on the state-owned oil company. Since a great part of its domestic expenditures (services, contractors, etc.) are not indexed at the OER and the foreign exchange sales to the Central Bank are indexed at the official exchange rate the company has incurred into important financial losses, therefore, needing assistance from the BCV and boosting inflation, as explained in section 4.

Given the increase of expenditures as a proportion of GDP, the exercise throws an average non-financial public sector loss of 7% of GDP during the period of study. During the legal period the loss reached 3% GDP points, while in the illegal period the loss mounted to 9% of GDP on average per year.

---

19 It is important to mention that when the parallel market was tolerated by the government Pdvsa was able to sell dollars at the parallel market rate.
LESSONS LEARNED
Theory and evidence supports that parallel exchange markets are not sustainable policies. Moreover, the toleration of a high-premium for long periods of time has non-desirable effects on resource allocation, inflation and growth. Therefore, there are no sound gains from maintaining the system in the long run. In this section, and by way of conclusion, I am going the sum-up the lessons learned from the Venezuelan experience and provide some arguments for the total overcoming of this policy.

First lesson: The system is misused and delivers terrible consequences
Parallel Markets should be a temporary option. Their effectiveness is inversely related to the size of premium and the length of time that it remains in place. They are just a way to smooth out the increase in the exchange rate to achieve the required real depreciation. Therefore, the depreciation of the parallel market rate just provides a safety valve to absorb the pressure coming from capital outflows, while OER works as a nominal anchor.

While PFXM seem to help control inflation and avoiding drops in real wages, this is not the case in the long run. In Venezuela, the short-run vision of adjustments shows that no clear efforts to correct monetary and fiscal policies in order to restore the macroeconomic balance affects the premium in the short run. The other evidence states that capital controls are used as an arrangement to maintain overvalued RER and expansionary macroeconomic policies in place for longer and therefore unsustainable periods. In this sense, the Venezuelan experience shows that this just worsens the picture: the country has exacerbated its Dutch disease condition and deteriorated its institutions, just as a consequence of the rising of the premium. This takes us to the second lesson.

Second lesson: The premium is everything
Theory and evidence have shown that macroeconomic fundamentals matter the most regarding premium determination. In particular, the premium reflects inconsistencies between policies that affect domestic absorption and the impossibility of maintaining the actual OER. Also, in the short run, the premium is largely affected by devaluation expectations of the official exchange rate, so creditability is a crucial issue. The Venezuelan evidence has clearly shown that the distinction between legal and illegal parallel exchange markets is not relevant regarding the determinants of the parallel exchange rate. Moreover, in the period of illegality the premium practically rose with an exponential trend.
Third lesson: Unification is a matter of commitment
A successful unification is one that can be sustained without leading to significant increases in inflation or recurrent BoP problems. Venezuela requires a sustainable exchange rate system. In countries with high fiscal deficits and inflation such as Venezuela, this means the adoption of crawling pegs or managed floats at a first stage. Unification to a fixed exchange rate fails or is more difficult because it needs a credible commitment to maintain external balance in the event of an external shock. In this case, strong fiscal balance and sufficient reserves (access to external credit) are key to withstand external shocks or speculation against the currency. Unifying without previously achieving those two is the chronicle of a death foretold.

In the end, it is not politics all about long term growth? Well, unification is important for long term growth. Large premiums create huge microeconomic distortions and induce corruption. Empirical evidence have shown that automatic access to foreign exchange for current and financial transactions creates large efficiency gains and increases transparency20.

Fourth lesson: Timing is key
In countries where parallel foreign exchange market was introduced to deal with capital flights, unification proceeded rapidly, generally as part of a comprehensive stabilization package. This was the case if Venezuela until 2010. Now, accumulated imbalances have made the country more alike the economies with extensive and long-lived controls. In this cases it could better to follow the African approach and advance to unification gradually.

Fifth lesson: The fiscal (and political) tail
The current exchange control in Venezuela has proved to be a case of macroeconomic and social failure. The system resulted in large premia, fiscal losses, smuggling, corruption scandals, (fiscal and human) capital losses, economic stagnation and chronic inflation. It seems, therefore, that a “rational” government would have pushed for unification years ago.

However, the fact that unification is a complicated process with a high risk of failure must not be an excuse after thirty years of economic loss. Therefore, the motivation for the use and maintenance of the system lies beyond

20 See World Bank, several reports.
simple economics. Political economy issues related to fears of devaluation, rent-seeking and clientelism seem probably to be at the end of the chain of the Venezuelan experience in the 21st Century. This paper calls for further research and proposals in this area.

REFERENCES


