Exchange Rate Fundamentals, Forecasting, and Speculation: Bayesian Models in Black Markets

Robert Gramacy

Booth School of Business, University of Chicago

Samuel W. Malone*

School of Management, University of the Andes, Bogotá, Colombia

Enrique ter Horst

IESA, Caracas, Venezuela

Abstract (Summary)

Although speculative activity is central to black markets for currency, the out-of-sample performance of structural models in those settings is unknown. We substantially update the literature on empirical determinants of black market rates and evaluate the out-of-sample performance of linear models and nonparametric Bayesian treed Gaussian process (BTGP) models against the random walk benchmark. Fundamentals-based models outperform the benchmark in out-of-sample prediction accuracy and trading rule profitability measures given future values of fundamentals. In simulated real-time trading exercises, however, the BTGP achieves superior realized profitability, accuracy, and market timing, while linear models do no better than a random walk.

JEL codes: F31, F37, G17

Keywords: Black markets, exchange rates, forecasting, arbitrage

*Corresponding author: University of the Andes School of Management, Cll. 21 No. 1-20 Piso 9, Bogotá, Colombia, Tel. +57 1 339 4949, ext. 3529; e-mail: samuel.malone@gmail.com; fax: +57 1 339 4949 ext. 2360

Acknowledgements: We would like to thank, without implication, Francis Diebold, Herman K. van Dijk, participants in the May, 2011 seminar at the Audencia School of Management in Nantes, France and participants at the November, 2011 European Seminar on Bayesian Econometrics (ESOBE) in Brussels, Belgium for their comments on this paper, and Neila Cáceres for her highly competent research assistance. All errors are our own. RBG would like to thank the Kemper Family Foundation for their support.

I. Introduction

In a seminal paper, Meese and Rogoff (1983) demonstrated that empirical models, in particular linear regression models and VARs, designed from the most important structural exchange rate models of the 1970s were inferior to a naïve random walk in out-of-sample forecasting performance, even when the realized next-period values of independent variables were taken as given. A great deal of literature, which we review selectively below, has repeatedly verified the basic conclusions of the Meese and Rogoff (1983) study, in particular for major cross-currency exchange rate pairs.

The same year in which Meese and Rogoff published their study, another seminal paper by Dornbusch et al. (1983) on "The Black Market for Dollars in Brazil" laid out the basic stock-flow model that has underlain most subsequent empirical work to date on the behavior of exchange rates in black markets for currency. However, a comprehensive evaluation of the out-of-sample performance of the black market structural models subsequently built upon the Dornbusch framework—the equivalent of the Meese and Rogoff exercise for black markets—has been entirely lacking from the literature. The absence of such a study is particularly puzzling in light of the widespread notion that speculative activity is a key driver of black market activity, and the fact that black markets, which possess certain special features that set them apart from regular currency markets, might be expected *a priori* to offer better opportunities for inter-temporal arbitrage to speculators with superior predictive abilities.

To fill this gap, the present paper exploits a monthly database of 34 black market episodes during the past fifty years to test the out-of-sample fit of structural black market models built from the variables that have proven most useful according to past studies of empirical regularities in black markets for currency. In fact, we go one step further: in addition to testing the out-of-sample fit of structural black market models when the next-period value of the structural variables is known, as in Meese and Rogoff (1983), we evaluate the success of trading rule profitability, predictive accuracy, and other measures when the next-period vector of fundamentals is unknown. The latter exercise provides a much more realistic evaluation of the ability (or lack thereof) of currency speculators to generate profits based on real-time trading. Against the standard benchmarks of the random walk with and without drift, we evaluate the performance of parametric linear regression models and a nonparametric Bayesian treed Gaussian process model due to Gramacy and Lee (2008), which makes no assumptions about the stationarity of the conditional relationship between exchange rates and fundamentals, or the stationarity of the exchange rate or fundamentals themselves. The adaptation of the BTGP model to exchange rate prediction, the first of its kind in the literature, is reasonable in light of previous findings, reviewed below, that indicate a role for nonlinear and non-stationary model features in improving exchange rate forecasting performance.

To preview the main findings of the paper, we find that in-sample, the implied monthly return to the "reverse carry trade" strategy of borrowing in (high interest) local currency and earning the (low interest) US interest rate in anticipation of large depreciations is by far the most robustly statistically significant determinant of the black market exchange rate across all 34 episodes studied.[1] The average coefficient across episodes for this variable, of 0.45, in our linear model in first differences implies that for a 1 percent increase in the local currency returns to the reverse carry trade strategy, the monthly depreciation rate of the black market currency relative to the US dollar increases by nearly half a percent, thus making this determinant economically very significant as well. The second most statistically significant determinant of the black market exchange rate across episodes is the official exchange rate, which is not surprising in light of the fact, emphasized by previous literature, that official devaluations typically have major impacts on black market exchange rates. Across episodes, a 100 percent official devaluation translates on average into a 59 percent contemporaneous depreciation of the black market exchange rate. Other variables emphasized in previous empirical studies of black market exchange rates, such as the logarithm of the ratio of M2 to the official exchange rate[2], the logarithm of the real official exchange rate (the official exchange rate multiplied by the ratio of US and domestic price indices), the growth rate of international reserves, and country-specific

[1] This variable, introduced under the name of the "differential rate of expected profits", or *DREP*, by Fishelson (1988) in his important early empirical study of black markets for dollars in a panel of 19 countries, measures the contemporaneous return in local currency terms to the reverse carry trade strategy described in the text. This variable can also be understood as measuring the realized deviation from uncovered interest rate parity (UIP).

[2] This variable, which proxies for the dollar value of local currency assets at the official exchange rate, affects the stock market equilibrium for black market dollars, as described in detail in Malone and ter Horst (2010), whose stock-flow model encompasses theoretical models proposed in the stock-flow tradition begun by the seminal paper of Dornbusch et al. (1983).

commodity price indices[3], are less robustly important across country episodes, although there are specific country episodes for which these variables are variously highly statistically significant at conventional levels.

Out-of-sample, all of our structural models significantly outperform the random walk and random walk with drift on measures of realized trading rule profitability in "ex-post" forecasting exercises in which the next period vector of fundamentals is given. Consistent with the findings of Meese and Rogoff (1983), however, we find that the random walk achieves significantly lower root-mean-squared-errors (RMSEs) and mean-absolute-errors (MAEs) than the structural models in this setting. The apparent conflict between the outperformance of structural models on measures of trading rule profitability and underperformance with respect to RMSE's is explained by the fact that the structural models achieve significantly better directional prediction accuracy, particularly in anticipation of large return events, despite the fact that their point predictions are further off from the true values than those of the random walk model on average. Our findings lend strong support to the claim that comparison of competing models based solely on their average errors and RMSE's may lead to very misleading conclusions with respect to their usefulness for the implementation of profitable speculative trading strategies. In "ex-ante" out-of-sample forecasting exercises, in which the next-period values of fundamentals is unknown, we find that while parametric linear models perform no better than a random walk, the BTGP model is capable of achieving economically and statistically significant trading profits—on the order of 17% annual percentage rate (APR) per month on average

[3] These last three determinants tested affect specifically the flow market equilibrium for black market dollars in the stock-flow model, as emphasized variously in the studies of Culbertson (1989), Phylaktis (1991), Shachmurove (1999), and Malone and ter Horst (2010).

across episodes. This suggests that the BTGP model may prove a useful forecasting engine in other financial market contexts as well, in particular foreign exchange.

The rest of this paper is organized as follows. Section II briefly summarizes the literature on empirical exchange rate models, with a focus on contributions stemming from the seminal Meese and Rogoff (1983) study. Section III lays out the basic empirical structural black market models, whose explanatory variables are inspired by contributions stemming from the Dornbusch et al. (1983) framework. Section IV describes our black market dataset. Section V reports in-sample results for the linear models, and substantially updates the literature on the empirical determinants of black market exchange rates by extending and comparing previous results in a common framework. Section VI describes the forecasting models, the trading strategy used in black markets based on the forecast signals, and the statistics used to measure model forecasting performance. Section VII reports results for the out-of-sample fit exercise, which is comparable with the exercises in the Meese and Rogoff tradition, as well as the out-of-sample forecasting exercise, in which the next-period values of fundamentals are unknown. Section VIII concludes.

II. Literature Review

In the years directly following the publication of Meese and Rogoff's (1983) paper, their findings were further ratified in work employing linear models with time varying features, such as Alexander and Thomas (1987), Wolff (1987), and Wolff (1988). Subsequently, given findings of nonlinear dependency in exchange rate changes (Baillie and McMahon (1989), Hsieh (1989) and Hong and Lee (2003)), some authors attempted to exploit such behavior to beat the forecasting accuracy of the random walk. Such attempts included Diebold and Nason (1990) and Meese and Rose (1990), who used non-parametric kernel regressions, and Engel and Hamilton (1990) and Engel (1994), who employed a Markov-switching model. None of these strategies proved useful in beating the random walk in out-of-sample forecasting performance.

More recently, Preminger and Franck (2007) report that work done on exchange rate forecasting using neural networks provides some evidence that they are better than other nonlinear models in terms of out-of-sample forecasting ability. The evidence on that score, however, is also mixed. While Kuan and Tung (1995), Brooks (1997) and Gencay (1999) all report that neural networks can beat a random walks for daily exchange rate data, for example by achieving lower root mean squared errors, a study by Qi and Wu (2003), who employ a neural network with monetary fundamentals at 1-month, 6-month, and 12-month horizons, finds that their model cannot beat the random walk model at those horizons. As Rogoff (2001) himself has noted, in a survey of the literature stemming from his paper with Meese, the inability of structural models to explain the movements of G3 exchange rates remains a robust stylized fact.

During the past decade, however, a handful of studies have managed to deliver more optimistic results with respect to the forecasting ability of structural models in foreign exchange. These studies have obtained their results, for the most part, either by implementing existing models using more sophisticated model selection criteria, or by expanding the range of exchange rates studied to include smaller country cross rates with the dollar or another major currency. Papers in the first group include Nag and Mitra (2002), Preminger and Franck (2007), and Sarno and Valente (2009). Nag and Mitra (2002) focus on potential improvements from model nonlinearity, providing evidence that a genetically optimized neural network model can modestly outperform GARCH type models on the three nominal exchange rate pairs examined in the Meese and Rogoff (1983) study. Preminger and Franck (2007) propose a robust regression approach, which is less sensitive to data contaminated by outliers, for improving the out-of-sample performance of the standard linear autoregressive and neural network models. Upon implementing their method for the pound/dollar and yen/dollar cross rates, they find that robust models tend to improve the forecasting ability of both types of models at all time horizons studied. Sarno and Valente (2009), using real-time data on a broad set of economic fundamentals for five major US dollar exchange rates, find that the difficulty of selecting the best predictive model is mostly due to frequent shifts in the set of fundamentals driving exchange rates, which can be interpreted as reflecting swings in market expectations over time. These authors employ a model selection procedure due to Pesaran and Timmermann (1995) to outperform a random walk for three out of five of the exchange rates studied. However, they also find that if conventional model selection criteria are used to choose the best model ex ante, the same set of economic fundamentals is not useful at all in forecasting exchange rates out-of-sample.

Papers in the second group, which expand the focus to smaller cross rates, include Liu, Gerlow, and Irwin (1994), Yang, Su, and Kolari (2008), and Cerra and Saxena (2010). Liu, Gerlow, and Irwin (1994) provide evidence showing that a monetary/asset model in a VAR representation does have forecasting value for some exchange rates. Yang, Su, and Kolari (2008) examine the potential martingale behavior of Euro exchange rates in the context of out-of-sample forecasts, with special attention paid to potential nonlinearity-in-mean. Their findings indicate that while martingale behavior cannot be rejected for Euro exchange rates with the yen, pound, and dollar, there is indeed nonlinear predictability in terms of economic criteria with respect to several smaller currencies. Most recently, Cerra and Saxena (2010) revisit the dynamic failure of the monetary models in explaining exchange rate movements by using information content from 98 countries to test for co-integration between nominal exchange rates and monetary fundamentals. They find robust evidence for co-integrating relationships, and that fundamentals-based models are very successful at beating a random walk in out-of-sample exchange rate prediction. The Cerra and Saxena paper, however, uses annual data on nominal exchange rates, many of which were pegged during the period of study, vis-à-vis the dollar.

While the above literature pertains to non-black market exchange rates, we believe that the main lessons of the papers discussed above that manage to achieve improved exchange rate forecasting performance can be capitalized on in the black market context as well. In particular, besides linear models, we employ a Bayesian treed Gaussian process (BTGP) model due to Gramacy and Lee (2008) to forecast one-month-ahead exchange rates. The BTGP is a nonparametric model that can be viewed as a tree whose leaves represent random functions, with the functions modeled using Gaussian processes. The flexibility afforded by being able to classify observations into different "leaves" of the BTGP allows for the modeling of non-stationary data, in which the function relating the black market rate and structural explanatory variables in a given month may represent one of several Gaussian processes indexed by the leaves of the tree. This is consistent with earlier findings that neural network models, and methods that explicitly allow for non-stationarity in the relationship between fundamentals and exchange rates, can improve out-of-sample prediction.

Second, besides the central fact that the out-of-sample performance of structural models is essentially unknown in the black market context, our focus on black markets is reasonable given the evidence that smaller cross-rates may provide a more fertile ground in which structural models may have a fighting chance against the random walk out-of-sample. Finally, in contrast to most previous studies, our focus is centered not only upon the question of whether or not out-of-sample forecasting ability for structural models exists, but also on whether or not such an ability confers the capacity to generate speculative profits in simulated real-time trading. For that reason, we go beyond the usual reporting of RMSE's and measures of out-of-sample fit when next-period values of fundamentals is known (à la Meese and Rogoff, 1983). We report directional accuracy measures, the Anatolyev-Gerko (2005) statistic of market timing ability, the cross-sectional distribution of realized monthly APR's from simulated trading, and contrast the results of the out-of-sample fit exercise with those of the out-of-sample forecasting exercise, in which next-period values of fundamentals is unknown, as it would be for currency speculators.

III. Structural Black Market Models

A. Linear Regression Models

The structural models developed in the analysis of black markets, which derive from the stock-flow model laid out by Dornbusch et al. (1983), use a different set of fundamentals as independent variables than the structural models of exchange rates tested by Meese and Rogoff (1983) and most subsequent studies. We have opted to focus directly in this paper on the fundamentals implied by the black market literature, rather than the set of variables used in studies of major exchange rate pairs based on the structural models of the 1970s.[4] Besides the fact that this approach is the most natural (and as the evidence will show, succeeds remarkably well) for the task of forecasting black market rates, we reserve the comparison of structural models developed for major exchange rates with black market structural models to future research.

The basic black market exchange rate regression in (log) levels takes the following general form:

$$e_t = \alpha + \beta^{\bar{e}}\bar{e}_t + \vec{\beta}'\vec{X}_t + \varepsilon_t \tag{1}$$

Here e_t is the log black market exchange rate, $\bar{e}_t$ is the log official exchange rate, $\vec{X}$ is a vector of structural variables taken from the black market literature, and ε_t is the error term. Since structural models in the tradition of Dornbusch et al. (1983) examine the black market premium, defined at the ratio of the black market exchange rate to the official exchange rate, and deliver a formula for the premium in terms of the elements of the vector of black market fundamentals $\vec{X}$, the only coefficient in equation (1) whose value is given by theory alone is $\beta^{\bar{e}} = 1$. This follows directly from log linearization, although the formulation of structural model of black market rates are consistent in this respect, and as we will see from our in-sample analysis, the value $\beta^{\bar{e}} = 1$ cannot be rejected as the mean estimate of $\hat{\beta}^{\bar{e}}$ in the cross section of black market episodes.

Articles that adapt Dornbusch et al. (1983) style stock-flow models of the black market into empirical models, and take those models to the data on the black market

[4] We refer, in particular, to the Frenkel-Bilson flexible price monetary model, the Dornbusch-Frankel sticky price monetary model, and Hooper-Morton sticky price asset model discussed and tested out-of-sample in Meese and Rogoff (1983).

premium, include Fishelson (1988), Culbertson (1989), Phylaktis (1991), Shachmurove (1999), and Malone and ter Horst (2010). Of these articles, only Malone and ter Horst (2010) evaluate the out-of-sample fit of their preferred regression model by computing the mean error, the RMSE, and the MAE. Using monthly data from the Venezuelan black market during the period from 2003 until 2009, they find that that a modified Fishelson (1988) style regression significantly outperforms an auto-regressive model in forecasting the black market premium out-of-sample when the ex-post values of explanatory variables are given, as in Meese and Rogoff (1983). They do not, however, compute statistical profitability measures for specific trading strategies, evaluate the prediction accuracy of their preferred model, or perform an out-of-sample forecasting exercise in which future values of fundamentals are unknown.

B. The Bayesian Treed Gaussian Process (BTGP) Model

The Bayesian treed Gaussian process (BTGP) is a nonparametric, nonstationary, nonlinear regression model designed with modeling flexibility and computational tractability in mind. It marries the smooth global perspective of an infinite basis expansion, via Gaussian processes (GPs), with the thrifty local adaptivity of partition models, via trees. GP regression is becoming widely recognized as the "default" nonlinear regression model. It has deep connections to many other nonparametric regression models, like radial basis function approximation, spline models, and neural networks (see, e.g., Rasmussen and Williams, 2006, who provide an excellent review), to name a few. Trees, on the other hand, are enjoying a resurgence in popularity due to recent demands from large-scale data mining applications. To help describe the BTGP we first provide a brief overview of its constituent parts.

GPs, which offer a flexible prior over functions, are uniquely and simply specified by a mean and covariance function. In the regression context this boils down to specifying the following multivariate normal relationship between a response n-vector *Y* and a design matrix of inputs *X*: $Y \sim N_n(f(X)\beta, \sigma^2 K)$, where $f(X)$ is a basis expansion of *X* that may be the identity or zero, possibly including an intercept or mean level; β are the usual linear regression parameters; and *K* is an *n* x *n* correlation (or covariance) matrix arising from a correlation (and noise) process that typically allows dependence between outputs to decay smoothly as a function of the distance between their covariates in *X*. An attractive feature of this setup is that inference for some of the parameters (e.g. β and σ^2) is analytic as in the linear regression context. Indeed, GP regression can be seen as a generalization of OLS. Inference for parameters to *K* requires numerical or Monte Carlo methods, but is not onerous.

The main feature of GP regression is that the predictive or forecasting equations are linear, available analytically given *K*. Forecasts thus obtained compare favorably to modern alternatives in out-of-sample validation comparisons. There are, however, two important drawbacks: computational tractability in large data sets, as an $O(n^3)$ operation is required to decompose *K* for inference and prediction; and stationarity of the predictive surface which is a typical simplifying assumption placed on *K*. Together this means that the data set being modeled must be small to moderate in size, and exhibit the same smoothness and noise characteristics throughout the input space. Such a barrier to entry may explain why GP regression has been slow to take off in some applied fields, such as empirical finance.

And this is where partitioning by trees comes in handy. Tree models have a rich history in computational statistics as a divide-and-conquer means of obtaining flexible regression and classification fits. Perhaps the most popular tree algorithm is CART (Classification and Regression Trees), by Brieman, et al. (1984). This algorithm uses a myriad of information theoretic criteria and cross-validation-like methods to learn a hierarchical patchwork of distinct regions of the input space, called the *leaves*, wherein simple regression or classification models can be fit. Since the leaves are fit independently, the aggregate surface can be highly nonlinear, exhibiting disparate spatial dependencies and thereby accommodate non-stationary relationships and heteroskedasticity. The abruptly changing predictive surfaces that result are visually, and sometimes theoretically, unattractive, when the application domain (e.g., fluid dynamics) dictates smoother relationships. The algorithm is fast and, despite such drawbacks, performs well in out-of-sample prediction problems in a wide array of applications.

A more recent advance in tree modeling involved specifying a prior over tree space so that likelihood-based methods of inference could be used (Chipman, et al., 1998). This allowed tree-space to be integrated over via Markov chain Monte Carlo (MCMC), resulting in a final predictive surface that evolves more slowly since it is not dependent on a single tree. It also helped facilitate the implementation of a wider array of more involved models at the leaf nodes. Gramacy & Lee (2008) showed how GPs could be employed at the leaves, resulting in a smoothly varying patchwork combining the best of both worlds: abrupt changes and smoothness as dictated by the relationships in the data. Their Bayesian treed Gaussian Process (BTGP) has since been applied successfully in areas as disparate as computational fluid dynamics, genetics, finance, climatology and political science. An important contributor to its success is the availability of open source software in the form on an R package called "tgp" available on CRAN (Gramacy, 2007).

IV. The Black Market Dataset

The structural models we consider in both our in-sample and out-of-sample analyses employ the same set of six independent variables, which have proven to be robust determinants of the black market premium in at least one of the past empirical studies. Our vector of fundamentals, $\vec{X}_t$, consists of the following variables: (a) the logarithm of the official exchange rate, $\bar{e}$; (b) the differential rate of expected profits, $DREP$, proposed by Fishelson (1988) in his empirical test of the model of Dornbusch et al. (1983) on 19 black market episodes; (c) the logarithm of the ratio of M2 to the official exchange rate, $\bar{c}$, (d) the logarithm of the real official exchange rate (the official exchange rate multiplied by the ratio of US and domestic price indices), $roer$; (e) the logarithm of the international reserves, r; and (f) the logarithm of the country's commodity price index, p, where the index is defined as the price of the country's primary export commodity, in US dollars, during the black market episode. The $DREP$ variable, which measures the deviation from uncovered interest rate parity, is calculated according to the formula $DREP_t = (1+i_t^*)(1+d_t) - (1+i_t)$, where i_t is the local currency monthly interest rate earned from time *t-1* to time *t*, $d_t = E_t / E_{t-1} - 1$ is the contemporaneous (time *t*) monthly rate of depreciation of the local currency relative to the US dollar, and i_t^* is the US monthly interest rate earned from time *t-1* to time *t*.[5] The practice of including the contemporaneous deviation from UIP, using either the official devaluation rate $\bar{d}_t$ or the black market depreciation rate d_t, in models of the contemporaneous black market exchange rate (or premium) is standard in the black market literature given the key assumption of the Dornbusch et al. (1983) model that this quantity should directly affect the equilibrium stock of black market dollars held by market participants. It is important to note that such a specification does not involve explaining the contemporaneous black market rate with itself, and that in fact the in-sample pairwise correlations between the log black market exchange rate e_t and the variable $DREP_t$ are usually small and positive, and frequently even negative, for most episodes in our sample.[6]

In the next section, as a preliminary to the out-of-sample analysis, we report in-sample results for the linear model (1) in levels and in differences for each of the 34 country episodes in our dataset. Besides updating previous work on empirical regularities in black markets for currency by Fishelson (1988), Culbertson (1989), and Shachmurove (1999), whose studies examined 19, 10, and 17 country episodes, respectively, our in-sample analysis substantially expands upon these previous multi-country studies both in terms of country coverage and the range of explanatory variables considered.

---

[5] Unlike Fishelson (1988), who was not able to find reliable local interest rate data in his study of 19 early black market episodes at the quarterly frequency and approximated those rates using inflation rates and the Fisher equation, we have sourced and used actual values for local interest rates at the monthly frequency for all episodes included in the present study.

[6] It is also worth nothing that in our linear model in differences, the log growth of the black market exchange rate is explained by the differenced *DREP* variable, rather than its level, and in our "ex-ante" forecasting exercises, in which the BTGP model succeeds in generating substantial out-of-sample trading profits before transactions costs, the *DREP* variable or its first difference, like other determinants, is lagged one month.

Our database of 34 black market episodes is constructed by supplementing data on parallel and official exchange rates provided by Reinhart and Rogoff (2004)[7] with data for Taiwan sourced from Luintel (2000) and the Taiwanese Central Bank, as well as data from Venezuela sourced from Malone and ter Horst (2010). Monthly data on macroeconomic aggregates was taken from the IMF and the US Federal Reserve, with country-specific commodity price indices sourced from the United Nations Commodity Trade Statistics. The first black market episode in our dataset began in January 1963 in India, several of the most recent episodes ended in July 1998, in the cases of South Africa, Paraguay, Jordan, and Egypt, and the last episode in our dataset belongs to Venezuela, whose policy of capital controls began in February of 2003, with our data on that country extending to July of 2009.

V. In-Sample Results for the Linear Structural Models

Given the breadth of our coverage of black market episodes in recent history, it is useful to document the performance of equation (1), in levels and in first differences, in-sample. This will provide an initial assessment of the ability of our benchmark parametric model to explain variations in the black market rate, plus an evaluation of the sign and significance of each determinant in driving changes in black market exchange rates. We include the results of the model in first differences due to the possibility that the black market rate or some subset of the determinants might be integrated of order 1. Tables OA.1a-OA.1c of the Online Appendix summarize the results of the in-sample exercise, performed at monthly frequency, for all of the countries in our sample.

7 Official home page: http://www.carmenreinhart.com/data/browse-by-topic/topics/10/

Overall, the six determinants selected from the black market literature perform well in the in-sample exercises in terms of their ability to explain the variations in the black market exchange rate. The F-statistics for joint significance are all significant at the 1% level, for the linear model in differences as well as in levels. The lowest value of the R-squared statistic for any country or model was 44%, for the Costa Rica episode for the linear model in levels, with the majority of models displaying R-squared statistics above 60%. While the models in levels tended to achieve higher R-squared values, they also tended to achieve higher RMSEs. The RMSEs for the model in differences range from 1%, in the cases of Costa Rica, Ireland, Jamaica, Malaysia, Paraguay, Philippines, Taiwan, Uruguay, and Colombia, to 13%, for the case of Uganda, with the majority of RMSEs being in the mid- to low-single digits. The high joint explanatory power of the six variables included in our linear models in-sample, and in-sample RMSEs that are lower, for a large fraction of countries, than the RMSEs achieved in the out-of-sample exercises of Meese and Rogoff (1983) for the random walk or structural models at the monthly frequency, raises the possibility that our structural black market model might stand a fighting chance against the random walk in the out-of-sample exercises we present in the following section.

[INSERT TABLE 1 ABOUT HERE]

Before proceeding to that analysis, and given the possibility that many of the level regressions above may be spurious due to first order integration of some variables, we performed an augmented Dickey-Fuller regression for each variable, in levels and differences, to test for a unit root. The results of those tests are presented in Table 1, and can be summarized as follows. With the exception of the variable DREP, both the black market rate and the other five explanatory variables in levels present strong evidence of first order integration. After taking first differences, however, the vast majority of (differenced) variables across episodes appear to be integrated of order zero. Thus, while an error correction model with a co-integrating equation might be a more efficient way of estimating the model in levels for most episodes, we can conclude that the t-statistics, F-statistics, and R-squared values obtained for the model in differences are valid for the vast majority of countries. For our purposes, this finding is sufficient to justify limiting the competitors to the random walk in out-of-sample fit and forecasting exercises to the linear model in levels and differences, plus the Bayesian treed Gaussian process model, as the latter is designed in a much more general way to detect and handle non-stationarity in the relationship between the dependent and explanatory variables.

[INSERT TABLE 2 ABOUT HERE]

The summary statistics of the point estimates of the linear model parameters, for the model in levels and in differences, are reported in Table 2. Several facts are worth mentioning. First, the null hypothesis $\beta^e = 1$, according to a two-sided t-test of the average value of this parameter given its variation across episodes for the model in levels, cannot be rejected at the 10% level. This is consistent with the theoretical value of unity for this parameter implied by the structural models of the black market literature. Second, the variable DREP is significant at the 10% level or better in 32 out of 34 episodes for the model in levels, and in all 34 of the episodes for the model in differences, with its average estimated value across countries remaining essentially the same, at around 0.45 (for every one percent increase in the realized deviation from uncovered interest rate parity, the black market rate increases by nearly half a percent), whether the model is estimated in levels or differences. This variable is the single most robust determinant of the black market exchange rate, consistent with the notion that speculative activity plays a key role in such markets. Finally, it should be noted that the incidence of episodes in which the other explanatory variables are significant falls when the linear model is run in differences, consistent with the previous finding that many of the variables in levels are integrated of order 1.

VI. Forecasting Models, the Black Market Trading Strategy, and Measurement of Results

Forecasting models

As a benchmark, inspired by Meese and Rogoff (1983), our first two forecasting models are: (1) a random walk, and (2) a random walk with drift, in which the drift is estimated using a rolling window of 12 months of data. The random walk with drift specification is reasonable to consider, as do Engel (1994) and Cerra and Saxena (2010), because the majority of black market episodes have occurred in the context of monetary expansions and display positive tendencies towards depreciation. The three structural models we employ are (3) a structural black market model in levels, (4) a structural black market model in differences, and (5) a Bayesian treed Gaussian process model in levels.[8]

We make a distinction, as in Howrey (1994), between "ex-post", or *out-of-sample fit* exercises, of the kind performed by Meese and Rogoff (1983), in which the vector of next-period fundamentals is given, and "ex-ante", or *out-of-sample forecasting* exercises, in which the next-period vector of fundamentals is unknown. For the random walk and random walk with drift, this distinction is moot, as only the time series information of the dependent variable is used for prediction. For the three structural models, however, this distinction is important. While the use of the next-period values of independent variables for prediction is standard in the exchange rate literature, and facilitates comparability with that literature, the out-of-sample forecasting exercises are far more relevant for currency speculators, who must take positions based only upon information available today. The models are as follows:

<u>Random walk:</u>

(Model 1) $e^f_{i,t+1} = e^f_{i,t}$

<u>Random walk with drift:</u>

(Model 2) $e^f_{i,t+1} = e^f_{i,t} + \hat{\alpha}_i$

<u>Structural Black Market Model in Levels:</u>

(Model 3 – "Ex-post" out-of-sample forecast equation)

$$e^f_{i,t+1} = \hat{\alpha}_i + \hat{\beta}^e_i e_{i,t+1} + \hat{\beta}^{DREP}_i DREP_{i,t+1} + \hat{\beta}^c_i c_{i,t+1} + \hat{\beta}^{roer}_i roer_{i,t+1} + \hat{\beta}^r_i r_{i,t+1} + \hat{\beta}^p_i p_{i,t+1}$$

[8] Our BTGP model results were obtained by running our R scripts for this model in the computer cluster of the Duke University Department of Statistics, which we would like to thank for this privilege.

(Model 3 – "Ex-ante" out-of-sample forecast equation)

$$e^{f}_{i,t+1} = \hat{\alpha}_i + \hat{\beta}^{e}_i e_{i,t} + \hat{\beta}^{DREP}_i DREP_{i,t} + \hat{\beta}^{c}_i c_{i,t} + \hat{\beta}^{roer}_i roer_{i,t} + \hat{\beta}^{r}_i r_{i,t} + \hat{\beta}^{p}_i p_{i,t}$$

Structural Black Market Model in Differences:

(Model 4 – "Ex-post" out-of-sample forecast equation)

$$e^{f}_{i,t+1} = e^{f}_{i,t} + \hat{\alpha}_i + \beta^{\Delta \bar{e}}_i \Delta \bar{e}_{i,t+1} + \hat{\beta}^{\Delta DREP}_i \Delta DREP_{i,t+1} + \hat{\beta}^{\Delta \bar{c}}_i \Delta \bar{c}_{i,t+1} + \hat{\beta}^{\Delta roer}_i \Delta roer_{i,t+1} + \hat{\beta}^{\Delta r}_i \Delta r_{i,t+1} + \hat{\beta}^{\Delta p}_i \Delta p_{i,t+1}$$

(Model 4 – "Ex-ante" out-of-sample forecast equation)

$$e^{f}_{i,t+1} = e^{f}_{i,t} + \hat{\alpha}_i + \beta^{\Delta \bar{e}}_i \Delta \bar{e}_{i,t} + \hat{\beta}^{\Delta DREP}_i \Delta DREP_{i,t} + \hat{\beta}^{\Delta \bar{c}}_i \Delta \bar{c}_{i,t} + \hat{\beta}^{\Delta roer}_i \Delta roer_{i,t} + \hat{\beta}^{\Delta r}_i \Delta r_{i,t} + \hat{\beta}^{\Delta p}_i \Delta p_{i,t}$$

Bayesian Treed Gaussian Process Model:

(Model 5 – "Ex-post" out-of-sample forecast equation)

$$e^{f}_{i,t+1} = f(\bar{e}_{i,t+1}, DREP_{i,t+1}, \bar{c}_{i,t+1}, roer_{i,t+1}, r_{i,t+1}, p_{i,t+1})$$

(Model 5 – "Ex-ante" out-of-sample forecast equation)

$$e^{f}_{i,t+1} = f(\bar{e}_{i,t}, DREP_{i,t}, \bar{c}_{i,t}, roer_{i,t}, r_{i,t}, p_{i,t})$$

Models (2)-(4) are estimated using rolling window of 12 months of data, rather than an expanding window, as this tended to produce better performance. The "ex-post" forecasting versions of the models are computed using the coefficients obtained in this manner, with the model predictions for time *t+1* computed by evaluating the model at the time *t+1* values of the vector of the explanatory variables. The "ex-ante" forecasting versions of the models estimate the coefficients for the model in which the explanatory variables are lagged one period, and the forecasts for period *t+1* are produced by evaluating the model at the period *t* values of the fundamentals.

As noted by Hendry and Mizon (2002) and Cerra and Saxena (2010), running the linear model in differences can have a smaller forecast bias than the same model in levels, because it is robust to forecasting after the equilibrium mean shift. Separate papers by Rossi (2005) and Flood and Rose (2007) have shown, in addition, that when the error term of the structural specification in equation (1) is highly serially correlated, the random walk can have a lower RMSE forecast even when the fundamentals have explanatory power. These issues are relevant for the structural model estimated in levels, and are mitigated by writing the model in growth rates (i.e. model (4) above), a specification that nests the random walk, as in Cerra and Saxena (2010). The lower in-sample RMSE values obtained in Section 4 for the linear model in differences further support this choice.

The "ex-post" (out-of-sample fit) and "ex-ante" (out-of-sample forecasting) versions of Model (5), the BTGP, are implemented in a way similar to the implementation of models (2)-(4), albeit with of two main differences. First, the BTGP is estimated using an expanding rather than a rolling window, given the fact that it is designed explicitly to handle non-stationary relationships between the dependent and structural variables through time, and as such is also more capable of making effective use of past information that would harm the performance of linear models due to such non-stationarity. Second, we estimate the BTGP at each month using two different prior distributions—the normal empirical Bayes prior and the independent mean zero normal prior with inverse-gamma variance—and then perform prediction using the out-of-sample fit/forecast estimate associated with the prior distribution that produces the corresponding predictive distribution for the log exchange rate with lowest variance. These prior distributions are described in more detail in Gelman et al. (2003). In general, when the signal-to-noise ratio is high, the empirical Bayes prior is most appropriate, and when the signal-to-noise ratio is low, the independent normal prior (or "proper" hierarchical prior, as it is sometimes called) is preferred for generating the point estimates of the next-period log exchange rate. The BTGP was run on the black market data in (log) levels only, rather than in differences.

As in Meese and Rogoff (1983), we report RMSE, MAE, and average error statistics for the logarithms of the actual and predicted exchange rates for each country episode. In addition, as suggested by Engel (1994), we report the proportion of forecasts that correctly predict the direction of change of the exchange rate. This metric is also reported by Cerra and Saxena (2010) for their out-of-sample exercises in testing the monetary model on annual panel data. In addition to these traditional measures of forecasting performance, we are also interested in the speculative profits that would have been attainable by speculators in the black markets we study.

To examine this issue, based on the forecasts of our five models, we employ a simple trading strategy inspired by the work of previous authors, in particular Anatolyev and Gerko (2005) and Gençay (1998). In the context of equity market speculation, these authors examine a strategy of buying shares of an index worth current wealth if the return forecast of the market index was greater than or equal to zero, and selling shares worth current wealth otherwise. The black market context of interest to us in the present paper requires an adaptation of this basic strategy, as follows.

The trading strategy

In the literature on foreign exchange, a popular and much analyzed trading strategy is the carry trade, which involves borrowing in a low interest rate currency and earning the local interest rate in a high interest rate currency during a specified time period, e.g. one month, while bearing the associated exchange rate risk (see e.g. Menkhoff et al., 2012 for a thorough discussion). Underlying the carry trade strategy is the assumption that the spot exchange rate follows a random walk, so that the forecast of the exchange rate return over the next period is zero, and the expected excess return from the carry trade strategy itself is simply equal to the interest rate differential.

To formulate a strategy based upon our model forecasts of the black market exchange rate return (or depreciation rate), we recognize that, unlike in the case of the random walk, the forecast of the depreciation rate conditional on fundamentals may be different from zero, and in particular, may exceed or fail to exceed the prevailing interest rate differential. This suggests a straightforward modification of the usual carry trade strategy: if the forecast for the depreciation rate of the local currency exceeds the interest rate differential, so that the expected excess return from the carry trade is negative, initiate a reverse carry trade, by borrowing in the high interest rate local currency and earning the US interest rate; on the contrary, if the forecast for the depreciation rate of the local currency is less than the interest rate differential, so that the expected excess return from the carry trade strategy is positive, then initiate a carry trade.

To be precise, the predicted excess US dollar return of the reverse carry trade is given by

$$R^{f}_{t+1|t} = 1 + i^{*}_{t+1} - (1 + i_{t+1})/(1 + d^{f}_{t+1|t}),$$

where $d^{f}_{t+1|t} = E^{f}_{t+1|t} / E_t - 1$ is the forecast for the rate of depreciation of the black market rate E_{t+1} given information at time *t*, i^{*}_{t+1} is the US interest rate on dollar deposits earned during period *t+1*, and i_{t+1} is the local currency interest rate earned during period *t+1*.[9] Our strategy is to initiate a reverse carry trade when $R^f_{t+1|t} \geq 0$, and to initiate a carry trade when $R^f_{t+1|t} < 0$. In other words, when the forecast depreciation rate $d^f_{t+1|t}$ of the currency is sufficiently high, so that we have

$$d^f_{t+1|t} \geq (1+i_{t+1})/(1+i_{t+1}{}^*)-1 \cong i_{t+1} - i_{t+1}{}^*,$$

we will execute a reverse carry trade during the period *t+1*, and otherwise, we will execute a carry trade.[10]

The EP, or *excess profitability*, statistic of Anatolyev and Gerko (2005) is based on a trading strategy that, given a forecast $R^f_{t+1|t}$ for the next-period returns of some asset, which in our case is the reverse carry trade, invests an amount equal to current wealth in the asset if $R^f_{t+1|t} \geq 0$, and shorts the asset in an amount equal to current wealth otherwise. The one-month return r_{t+1} from this trading strategy is equal to $r_{t+1} = sign(R^f_{t+1|t})R_{t+1}$, where the realized strategy return R_{t+1} is given by

$$R_{t+1} = 1 + i^*_{t+1} - (1+i_{t+1})/(1+d_{t+1}),$$

[9] Note that i^*_{t+1} and i_{t+1} are (in a slight abuse of notation, but consistent with our STATA code) technically the only variables whose values at time *t+1* are known in advance at the end of period *t*, and are thus in the information set at time *t*.

[10] Note that, as we have formulated our forecasting models in terms of the log exchange rate, the term $1+d^f_{t+1|t} = E^f_{t+1|t}/E_t$ in the denominator of the expression for $R^f_{t+1|t}$ is equal to $\exp(e^f_{t+1|t} - e_t)$, the exponentiated log return forecast. We, like previous literature in the Meese and Rogoff (1983) tradition, formulate our forecasting models in terms of the log exchange rate, in order to avoid Siegel's paradox, which dictates that unbiased forecasts of $E^f_{t+1|t}$ will always lead to biased forecasts of $1/E^f_{t+1|t}$, and vice versa.

and $d_{t+1} = E_{t+1} / E_t - 1$ is the realized rate of currency depreciation. For realized returns R_{t+1} and return forecast $R^f_{t+1|t}$, the Anatolyev-Gerko statistic is computed as

$$EP = \frac{A_T - B_T}{\sqrt{\hat{V}_{EP}}} \xrightarrow{d} N(0,1),$$

where $$A_T = \frac{1}{T}\sum_{t=1}^{T} sign(R^f_{t|t-1})R_t \text{ and } B_T = \left(\frac{1}{T}\sum_{t=1}^{T} sign(R^f_{t|t-1})\right)\left(\frac{1}{T}\sum_{t=1}^{T} R_t\right),$$

T is the number of out-of-sample periods, and $\hat{V}_{EP}$ is the feasible estimate of the variance of the numerator of the statistic, given in Anatolyev and Gerko (2005).

The EP statistic is useful because is measures the extent to which a given forecasting signal is able to generate trades whose profitability (before transactions costs) exceeds that of trades generated by a forecasting signal whose unconditional directional forecasting probabilities are equal to those of the model being tested. Models with significantly positive EP statistics, therefore, are often able to achieve a consistent market timing of nontrivial moves of the exchange rate. We report the EP statistic of Anatolyev and Gerko (2005), as well as our other statistics, for both the “ex-post” out-of-sample and “ex-ante” out-of-sample forecasting exercises.

In the context of “ex-post” forecasts, in which the ex-post values of the independent variables are given, a positive and significant EP statistic indicates superior market timing ability, conditional upon perfect foresight of the structural model determinants one month ahead. In the context of our “ex-ante” out-of-sample forecasting exercise, in which future values of fundamentals are unknown, a positive EP statistic suggests a potentially profitable superiority in market timing ability. Our application of the Anatolyev-Gerko EP statistic to the measurement of market timing ability in the context of foreign exchange, via the formulation of the trading strategy described above, is to our knowledge the first in the literature, and likely to be of interest to others wishing to measure the predictive ability of empirical exchange rate models in more conventional foreign exchange markets.

An additional measure of results

Finally, as a transparent measure of realized profitability, we compute the cumulative monthly return, expressed as an APR, for each model and country in the sample. In all of our exercises, unless otherwise stated, the initial training window is set equal to the first 12 months of data, and the out-of-sample testing period comprises the last $T_j - 12$ months of data, where T_j is the total number of months for episode *j*, with country episodes in alphabetical order by country indexed by $1 \le j \le 34$. Thus the monthly APR for country *j*, assuming an initial capital of USD 1000, is computed as $APR_j = 12((V_{T_j} / 1000)^{1/(T_j - 12)} - 1)$, where V_{T_j} is the value of the portfolio after the final month of the black market period.

VII. Results on "Ex-post" and "Ex-ante" Out-of-Sample Forecasting Exercises

A couple of points are worth mentioning as a preface to reporting the results. In a few exceptional cases, the BTGP needed additional months in the training data phase to converge in comparison to the linear models. In the case of Jamaica, 17 months were needed instead of 12, and in the case of Egypt, substantially many more months were needed for the algorithm to converge. We opted to leave Jamaica in the sample, with 43 versus 48 out-of-sample data points, and to remove Egypt entirely from the sample averages for model 5. Also, in a few exceptional cases, bankruptcy was arrived at by at least one of the models during the out-of-sample tests. These cases included, for the "ex-post" forecasting exercise: Morocco and Myanmar for the random walk and random walk with drift (models 1 and 2, respectively), and Morocco for the BTGP (model 5). For the "ex-ante" forecasting exercise, cases of bankruptcy included Morocco and Myanmar for both the random walk and random walk with drift, Myanmar for the linear model in differences (model 4), and Morocco for the BTGP. In cases of bankruptcy, AE, RMSE, and MAE statistics may still be computed in the same manner, as can the Anatolyev-Gerko statistic, the directional prediction accuracy percentages, and the Theil U-statistic versus the benchmark. Thus, in the exceptional cases of bankruptcy, we opted to include these countries in the calculations of the aforementioned statistics, and omit them from the calculations of the summary statistics of the monthly APR's by episode, as the monthly APR in cases of bankruptcy is not well defined.

Table 3 contains summary statistics for the ME, RMSE, and MAE calculations, both in the case of the "ex-post" out-of-sample fit exercise (panel A) and the "ex-ante" out-of-sample forecasting exercise (panel B). Table 4 displays the summary statistics, across our sample of 34 episodes, for realized cumulative monthly returns, Theil U-statistics (specifically, instances of significance at the 5% level for either the benchmark or alternative model), the proportion of correct directional forecasts, and the Anatolyev-Gerko results (instances of significance at the 5% level along with the sign of the EP statistic), with out-of-sample fit and forecasting exercises, respectively, reported in panels A and B of the table.

[INSERT TABLE 3 ABOUT HERE]

Let us turn first to the results displayed in Table 3. For both the “ex-post” (panel A) and “ex-ante” forecasting (panel B) exercises, all models appear to achieve average errors close to zero, with the structural models even slightly outperforming the random walk and random walk with drift on that measure. The RMSE’s, however, are strictly increasing by model number for both exercises, with the BTGP displaying the highest RMSE’s in both panels A and B, and attaining RMSE’s that are nearly three times those achieved by the random walk. The MAE’s, in both panels A and B, follow essentially the same pattern. Not surprisingly, the RMSE’s and MAE’s for nearly all models are slightly lower in panel A than in panel B, owing to the extra uncertainty introduced in the “ex-ante” exercise where the next-period values of the fundamentals are unknown. It is instructive to compare the results displayed in panel A of Table 4 with the RMSE’s reported in Table 1 of Meese and Rogoff (1983) for major currency cross rates and structural exchange rate models of the 1970s. Compared to the monthly RMSE’s reported by those authors, there are two main differences of note. First, the average RMSE obtained for the random walk is just over 8%, whereas the comparable figure for the yen, mark, and pound was just over 3% on average in the Meese and Rogoff (1983) study. This is not surprising, and reflects the higher levels of volatility observed in black markets for currency. Second, the structural models studied by Meese and Rogoff (1983) exhibit monthly RMSE’s that are only slightly higher than those of the random walk in that setting, with the RMSE obtained from the VAR model approximately double that of the linear structural models, at around 6-7% on average. In our case, linear structural black market models in levels and differences exhibit RMSE’s that are 50-60% higher than the benchmark RMSE's obtained by the random walk (12-13% versus 8%). The RMSE of the BTGP is even higher, at around 22%. These differences, which may be explained by the use of a wider range of fundamentals in our exercise, and higher levels of fundamental volatility, nevertheless are instructive, as they suggest that the underperformance (with respect to RMSE's) of fundamental models in the black market context is even starker than in the setting examined by Meese and Rogoff (1983) and many subsequent authors.

An examination of Table 4, nonetheless, suggests that assessing relative model performance according to the RMSE criterion may be very misleading when it comes to the issue of the potential to generate speculative profits. In particular, although the tabulation of Theil U-statistics of the structural models versus the benchmarks confirms clear outperformance of the random walk (with and without drift) on the RMSE criterion, a look at the proportion of correct directional guesses tells a quite different story. In the "ex-post" out-of-sample fit exercise (panel A), the linear structural model in differences achieves an average directional accuracy across episodes of 70.1%, versus only 57.6% for the random walk. The BTGP does slightly better than the random walk, with a directional accuracy of 60.8%. When we turn to the realized monthly APRs generated by each model, the dominance of structural models is clear and quite impressive: conditional on perfect foresight of the next-period values of fundamentals, the average realized APRs across episodes generated by our simple trading strategy follow the order of the directional accuracy measures exactly, and range from 35.7% for the linear model in differences, to -5.5% for the random walk. The BTGP is in the middle of the pack, with an average monthly return of 17.6% APR across episodes. The average monthly APRs for the three structural models are all significantly different from zero in the cross-section of episodes at the 1% level, as each average structural model error is greater than four times the cross-sectional standard error for the APR. Neither the random walk nor the random walk with drift, however, produces a mean trading strategy return that is significantly different from zero across episodes.

Turning to panel B of Table 4, we see that in the "ex-ante" out-of-sample forecasting exercise, as in the "ex-post" exercise in panel A, the Theil U-statistic confirms statistically significant outperformance of the random walk when it comes to generating lower RMSE's versus the structural models. Also, the average directional accuracy statistics reveal that the two linear models achieve slightly lower correct guessing percentages than the random walk and random walk with drift—54.1% and 53.4% for the linear model in levels and changes, respectively, versus 57.6% for the random walk and 56.6% for the random walk with drift. The BTGP achieves the highest directional accuracy of all, at 58.8%. A large difference emerges, in addition, between the average monthly APR's of the structural strategies, with both linear strategies earning returns that are statistically indistinguishable from zero, while the BTGP manages to earn an average monthly APR across episodes, before transactions costs, of 17%. The latter figure is the only return that is statistically different from zero in the cross-section, and at 3.9 standard errors from zero, is significant at the 1% level. Surprisingly, the average realized trading strategy returns for the BTGP in panel B, in which the next-period values of fundamentals was unknown, is only slightly lower than the average value across episodes achieved in panel A, when next-period values were given. It is clear that, even with substantial transactions costs, for example on the order of 5% APR per month, the realized excess trading strategy returns for the BTGP displayed in panel B would remain statistically and economically very significant.

[INSERT TABLE 4 ABOUT HERE]

Finally, recall that the Anatolyev-Gerko EP statistic measures the market timing ability of the strategy being tested against the market timing ability of a benchmark strategy whose directional guessing accuracy is equal to that of the strategy being tested, but which forecasts the direction of the dependent variable according to the flip of a coin biased according to the aforementioned directional probabilities. While the EP (excess profitability) statistic is not a direct measure of profitability *per se*, since the benchmark "naïve" strategy may or may not be profitable (and transactions costs have not been taken into account), it is a good measure of the ability of the strategy being tested to correctly identify the precise periods in which the dependent variable will rise or fall. In the out-of-sample fit exercise in panel A of Table 4, the linear model in levels obtains an EP statistic that is positive and significant at the 5% level for a total of 16 countries, the most of any model, followed by the random walk with 13, the linear model in differences with 12, the BTGP with 11, and the random walk with drift with 2. The three structural models can clearly be said to have outperformed the random walk models, however, as they generate far fewer instances of negative EP statistics that are significant at the 5% level, versus 20 such cases for the random walk, and 27 for the random walk with drift.

In panel B of Table 4, both linear models perform substantially worse on the EP score compared to their performance in the out-of-sample fit exercise, with both the linear model in levels and in differences obtaining only 2 instances of positive EP statistics significant at the 5% level, and 17 and 15 instances of negative EP statistics significant at the 5% level, respectively. The BTGP, however, slightly improves its performance in the "ex-ante" out-of-sample forecasting exercise, obtaining 13 instances of positive EP statistics significant at the 5% level, and only 9 instances of negative EP statistics significant at the 5% level. On the whole, the results of the EP test suggest that, in out-of-sample fit exercises where future values of fundamentals are given, linear structural models exhibit somewhat better market timing ability compared to the random walk benchmarks, and results largely comparable to those of the BTGP. In the simulated real-time trading exercises most relevant to speculators, however, the BTGP model exhibits a market timing ability superior to that of all other models, although again in this context the most naïve model, the random walk, clearly dominates the linear structural models.

There are perhaps three major lessons to take away from the results just presented. First, RMSE results alone may be very misleading when it comes to evaluating the economic value of model forecasts when it comes to the ability to generate speculative profits. Second, in out-of-sample fit exercises, linear structural models outperform both the BTGP model and the benchmarks in terms of accuracy and realized profitability. Third, when next-period values of fundamentals are unknown, as is the case for real-world speculators, the BTGP model outperforms all other models, and linear structural models are unable to produce profits that are statistically different from zero or from the benchmarks. It can be inferred that the additional uncertainty involved in forecasting next-period fundamentals is what separates, to a significant extent, the BTGP model from the linear models, as the more general structure of the BTGP allows it to capture the evolution of the joint distribution of the fundamentals simultaneously with the evolution of the black market exchange rate.

A note on real world profitability: transactions costs in black markets

The average APR of 17% across black market episodes found for the simple monthly trading strategy based on the predictions from the Bayesian treed Gaussian process is highly economically significant, and it bears repeating that this figure does not take into account the bid-ask spread present in the markets studied, which is likely to be substantially greater than bid-ask spreads typically found in large, liquid foreign exchange markets.[11] Although very little data exists on the bid-ask spread in black markets, one exception is an early paper by Dornbusch and Pechman (1985), who examine the bid-ask spread in the black market for dollars in Brazil. In terms of monthly averages of daily data extracted from the *Journal do Comercio*, they report that the bid-ask spread in that market during the period from March 1979 to December 1983 ranged from a low of 1.9 percent to a high of 8.4 percent, with a mean of 3.6 percent. Based on these figures, it is plain to see that our (reverse) carry trade strategy would not be profitable on average for a typical market participant. As an example, the approximately 1.5 percent average monthly profit we find prior to transactions costs would turn into an approximately 2.2 percent loss net of transactions costs for a typical carry trade involving borrowing at a 0.5 percent monthly interest rate in dollars, earning a 2 percent interest rate in local currency, and assuming a realized monthly depreciation rate of zero, with the bid set at 96.4 percent of the ask rate quoted by the local dealer, consistent with the average monthly spread reported in the

[11] We would like to thank an anonymous referee for emphasizing this point.

Dornbusch and Pechman (1985) study.[12] However, it should also be noted that sophisticated local dealers capable of employing our strategy, in the face of lower transactions costs given their natural position as a market maker, might indeed be able to reap economically significant profits net of the transactions costs they face in such contexts. Also, it is not clear to what extent we can generalize the findings on the bid-ask spread reported in the Dornbusch and Pechman (1985) study to other black market episodes, especially the more recent ones, in our sample. What is certain is that the level and drivers of transactions costs in black markets for currency appears to be a sorely understudied subject worthy of further attention.

VIII. Conclusion

In this paper we have substantially updated the literature on the drivers of black market exchange rates, showing that the official rate and the realized deviations from uncovered interest rate parity (UIP) are by far the most robust determinants of the black market rate across all 34 episodes studied. Further, we have shown that, while structural models of black market exchange rates tend to underperform the random walk with respect to average RMSE's in out-of-sample fit exercises, they exhibit superior directional accuracy, and this translates into a substantial ability to generate speculative profits, prior to transaction costs, conditional upon perfect foresight of next-month values of fundamentals. In out-of-sample forecasting exercises where futures values of fundamentals is unknown, as would be the case for real-world currency speculators, linear structural models underperform the random walk in nearly every test, from RMSE's and directional accuracy, on the one hand, to the Anatolyev-Gerko (2005) test of market timing ability and average

[12] Carry Trade Return = 1.02*(96.4/100) - 1.005 = -0.022

realized profitability, on the other. The Bayesian treed Gaussian process (BTGP) model of Gramacy and Lee (2008), however, exhibits superior market timing ability, directional accuracy, and realized profitability in simulated real-time trading versus the random walk and all other models, despite having significantly worse RMSE's than the random walk. While uncertainty about future fundamentals, and a potentially non-stationary relationship between fundamentals and black market exchange rates prevents linear structural models from converting their successful out-of-sample fit into profitable out-of-sample forecasting performance, the BTGP manages to achieve average realized monthly profits prior to transaction costs of 17% APR, a figure which is significantly different from zero at the 1% level in the cross-section. Future research is needed to evaluate whether the success of the BTGP in out-of-sample exchange rate forecasting based on structural models can be obtained to the same degree in regular markets for foreign exchange, and if so, what impact this has on the economic value that can be extracted from empirical exchange rate models (as in Abhyankar, Sarno and Valente, 2005 and Della Corte, Sarno, and Tsiakas, 2009).

References

Abhyankar A, Sarno L, Valente G. 2005. Exchange rates and fundamentals: Evidence on the economic value of predictability. *Journal of International Economics* **66(**2): 325-348. DOI: 10.1016/j.jinteco.2004.09.003.

Alexander D, Thomas III LR. 1987. Monetary/asset models of exchange rate determination. How well have they performed in the 1980's? *International Journal of Forecasting* **3**(1): 53-64.

Anatolyev S, Gerko A. 2005. A Trading Approach to Testing for Predictability. *Journal of Business and Economic Statistics* **23**(4): 455-461. DOI: 10.1198/073500104000000640.

Baille RT, McMahon PC. 1989. *The Foreign Exchange Market: Theory and Econometric Evidence*. New York: Cambridge University Press.

Breiman L, Friedman JH, Olshen R., Stone C. 1984. *Classification and Regression Trees*. Wadsworth: Belmont, CA.

Brooks C. 1997. Linear and nonlinear (non-)forecastability of daily Sterling exchange rates. *Journal of Forecasting* **16**(2): 125–145.

Cerra V, Saxena SC. 2010. The monetary model strikes back: Evidence from the world. *Journal of International Economics* **81(**2): 184-196. DOI: 10.1016/j.jinteco.2010.03.003.

Chipman, H.A., George, E.I., and McCulloch, R.R., 1998. "Bayesian CART model search (with discussion)", *Journal of the American Statistical Association*, 93: 935-960.

Culbertson WP. 1989. Empirical regularities in black markets for currency. *World Development* **17**(12): 1907-1919.

Della Corte P, Sarno L, Tsiakas I. 2009. An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies* **22**(9): 3491-3530. DOI: 10.1093/rfs/hhn058

Diebold FX, Nason JA. 1990. Nonparametric exchange rate prediction? *Journal of International Economics* **28**(3-4): 315–332.

Dornbusch R, Dantas DV, Pechman C, De Rezende Rocha R, Simões D. 1983. The Black Market for Dollars in Brazil. *Quarterly Journal of Economics* **98**(1): 25-40.

Dornbush, R and Pechman, C, "The Bid-Ask Spread in the Black Market for Dollars in Brazil," *Journal of Money, Credit, and Banking*, November 1985: 517-520.

Engel C. 1994. Can the Markov switching model forecast exchange rates? *Journal of International Economics* **36**(1-2): 151-165.

Engel C, Hamilton JD. 1990. Long swings in the dollar: Are they in the data and does the market know it? *American Economic Review* **80**(4): 689–713.

Fishelson G. 1988. The black market for foreign exchange: An international comparison. *Economics Letters* **27**(1): 67-71.

Flood RP, Rose AK. 2007. Why so Glum? The Meese-Rogoff Methodology Meets the Stock Market. *Centre for Economic Policy Research Discussion* Papers no. DP6714.

Gençay R. 1999. Linear and nonlinear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics* **47**(1): 91–107.

Gençay R. 1998. Optimization of technical trading strategies and the profitability in security markets. *Economics Letters* **59**: 249-254.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. *Bayesian Data Analysis,* Second Edition, London: CRC Press.

Gramacy RB. 2007. tgp: An R Package for Bayesian Nonstationary, Semiparametric Nonlinear Regression and Design by Treed Gaussian Process Models. *Journal of Statistical Software* **19**(9).

Gramacy RB, Lee HKH. 2008. Bayesian Treed Gaussian Process Models With an Application to Computer Modeling. *Journal of the American Statistical Association* **103**(483): 1119-1130.

Hendry D, Mizon GE. 2002. Forecasting in the Presence of Structural Breaks and Policy Regime Shifts. Working Paper. Nuffield College: University of Oxford.

Hong Y, Lee TH. Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics* **85**(4): 1048–1062.

Howrey EP. 1994. Exchange rate forecasts with the Michigan quarterly econometric model of the US economy. *Journal of Banking & Finance* **18**(1): 27-41.

Hsieh DA. 1989. Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business* **62**(3): 339-368.

Kuan C-M, Tung L. 1995. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics* **10**(4): 347–364.

Liu TR, Gerlow ME, Irwin SH. 1994. The performance of alternative VAR models in forecasting exchange rates. *International Journal of Forecasting* **10**(3): 419-433.

Luintel KB. 2000. Real Exchange Rate Behaviour: Evidence From Black Markets. *Journal of Applied Econometrics* **15**(2): 161-185.

Malone SW, ter Horst E. 2010. The Black Market for Dollars in Venezuela. *Emerging Markets Finance and Trade* **46**(5): 71-93. DOI: 10.2753/REE1540-496X460505.

Meese RA, Rogoff K. 1983.Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* **14**(1-2): 3-24.

Meese RA, Rose AK. 1990. Non-linear, non-parametric, non-essential exchange rate estimation. *American Economic Review* **80**(2): 192–196.

Menkhoff, L., Sarno, L., Schmeling, M. and A. Schrimpf (2012). "Carry Trades and Global Foreign Exchange Volatility", *The Journal of Finance*, Vol. 67 (2): 681-718.

Nag AK, Mitra A. Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting* **21**(7): 501-511. DOI: 10.1002/for.838.

Pesaran MH, Timmermann A. 1995. Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance* **50**(4): 1201-1228.

Phylaktis K. 1991. The Black Market for Dollars in Chile. *Journal of Development Economics* **37**(1–2): 155–172.

Preminger A, Franck R. 2007. Forecasting exchange rates: A robust regression approach. *International Journal of Forecasting* **23**(1): 71-84. DOI: 10.1016/j.ijforecast.2006.04.009

Qi M, Wu Y. 2003. Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* **10**(5): 623–640. DOI: 10.1016/S0927-5398(03)00008-2

Rasmussen CE, Williams CKI. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.

Reinhart CM, Rogoff KS. The Modern History of Exchange Rate Arrangements: A Reinterpretation. *Quarterly Journal of Economics* **119**(1): 1-48. DOI: 10.1162/003355304772839515

Rogoff, KS. 2001. The Failure of Empirical Exchange Rate Models: No Longer New, But Still True. *Economic Policy*, web essay at http://www.economic-policy.org/pdfs/responses/Kenneth-Rogoff.pdf

Rossi B. 2005. Testing Long-horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle. *International Economic Review* **46**(1): 61–92. DOI: 10.1111/j.0020-6598.2005.00310.x.

Sarno L, Valente G. 2009. Exchange Rates and Fundamentals: Footloose or Evolving Relationship? *Journal of the European Economic Association* **7**(4): 786-830.

Shachmurove Y. 1999. The Premium in Black Foreign Exchange Markets: Evidence from Developing Economies. *Journal of Policy Modeling* **21**(1): 1-39.

Yang J, Su X, Kolari JW. Do Euro exchange rates follow a martingale? Some out-of-sample evidence. *Journal of Banking and Finance* **32**(5): 729-740. DOI: 10.1016/j.jbankfin.2007.05.009

Wolff, CCP. 1987. Time Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models. *Journal of Business & Economic Statistics* **5**(1): 87–97.

Wolff, CCP. 1988. Models of exchange rates: A comparison of forecasting results. *International Journal of Forecasting* **4**(4): 605–607.

TABLES (Tables 1, 2, 3, and 4)

Table 1. Average augmented Dickey-Fuller test statistics and p-values across episodes for variables in levels and differences

	Levels						
	e	$\bar{e}$	$DREP$	$\bar{c}$	$roer$	r	p
Avg. ADF test statistic	-1.528617	-1.128112	-7.836299	-1.684567	-1.766066	-1.858649	-1.393658
Avg. p-value	.5511364	.6572991	.0007091	.505599	.5103677	.4721053	.5671836
	Differences						
	Δe	$\Delta \bar{e}$	$\Delta DREP$	$\Delta \bar{c}$	$\Delta roer$	Δr	Δp
Avg. ADF test statistic	-8.11247	-6.480458	-12.67555	-12.14543	-6.820205	-7.613862	-5.954635
Avg. p-value	.0008661	.0404112	6.50e-07	.0020095	.0067459	.0029218	.0111197

***** Table 1 displays the sample averages, across the 34 black market episodes, of the augmented Dickey-Full test statistics and associated p-values for each variable listed. Results are shown for the variables in levels (top panel) and first differences (bottom panel). The ADF test equation and null hypothesis for the ADF test are shown below, where *x* represents the variable being tested:

ADF test equation: $\Delta x_t = \alpha + \gamma x_{t-1} + \delta_1 \Delta x_{t-1} + \varepsilon_t$

Null hypothesis: $\gamma = 0$

Table 2. Summary statistics of point estimates for model coefficients across episodes

	Linear model in levels						
	α	$\beta^{\bar{e}}$	β^{DREP}	$\beta^{\bar{c}}$	β^{roer}	β^{r}	β^{p}
Mean	.7447638	1.005669	.4636094	.0626091	-.1067344	-.0548011	.0728697
St. Dev.	7.739526	1.273033	.1585501	.4293493	.6990775	.1691361	1.569532
Min	-32.13485	-1.212451	.0210541	-.6696389	-1.56892	-.7680858	-2.758923
Max	14.0946	7.09862	.8783977	1.915627	1.937548	.1662522	7.231605
Incidence of significance at the 10% level	22	30	32	19	21	22	22
	Linear model in differences						
	α	$\beta^{\Delta\bar{e}}$	$\beta^{\Delta DREP}$	$\beta^{\Delta\bar{c}}$	$\beta^{\Delta roer}$	$\beta^{\Delta r}$	$\beta^{\Delta p}$
Mean	.0064986	.5899542	.4509227	-.0118268	-.0545028	-.0260724	.023007
St. Dev.	.0156786	.7522906	.0916729	.2237978	.4269505	.0829999	1.3197
Min	-.0231931	-.699896	.0428525	-.7456525	-1.318188	-.4223309	-5.85733
Max	.0580682	3.786434	.5256265	.6747291	.8715442	.0668854	4.710825
Incidence of significance at the 10% level	6	14	34	6	7	4	2

* Table 2 reports the mean, standard deviation, minimum value, maximum value, and incidence of statistical significance at the 10% level or better of the coefficients for the variables shown, across the 34 episodes, for the linear model in levels (top panel) and differences (bottom panel). These summary statistics are drawn from the results on the point estimates for the coefficients reported in Tables OA.1a-OA.1c. The linear models in levels and differences are reproduced below for convenience.

Linear model in levels: $e_t = \alpha + \beta^{\bar{e}}\bar{e}_t + \vec{\beta}'\vec{X}_t + \varepsilon_t$

Linear models in differences: $\Delta e_t = \alpha + \beta^{\bar{e}}\Delta\bar{e}_t + \vec{\beta}'\Delta\vec{X}_t + \varepsilon_t$

Table 3: Sample error statistics for out-of-sample fit and forecasting exercises, models (1) – (5).

	Model:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	Number of Countries	34	34	34	34	33
Panel A. Out-of-sample Fit (Ex-post values of independent variables known)						
Mean Error	Mean:	-0.0163	-0.0011	-0.0078	0.0039	-0.0059
	St. Dev.:	0.0341	0.0081	0.0403	0.0323	0.0166
	Min:	-0.1646	-0.0419	-0.2202	-0.0544	-0.0440
	Max:	0.0060	0.0141	0.0205	0.1343	0.0387
RMSE	Mean:	0.0819	0.0821	0.1258	0.1360	0.2230
	St. Dev.:	0.0752	0.0736	0.1206	0.1490	0.1776
	Min:	0.0173	0.0173	0.0170	0.0152	0.0265
	Max:	0.2697	0.2806	0.6111	0.7079	0.7289
MAE	Mean:	0.0545	0.0554	0.0757	0.0667	0.1473
	St. Dev.:	0.0528	0.0507	0.0716	0.0642	0.1094
	Min:	0.0100	0.0105	0.0118	0.0102	0.0168
	Max:	0.2012	0.2003	0.3327	0.2883	0.4473
Panel B. Out-of-Sample Forecasting (Ex-post values of independent variables unknown)						
Mean Error	Mean:	-0.0163	-0.0011	-0.0067	-0.0022	-0.0065
	St. Dev.:	0.0341	0.0081	0.0231	0.0398	0.0215
	Min:	-0.1646	-0.0419	-0.0949	-0.1560	-0.0553
	Max:	0.0060	0.0141	0.0362	0.1236	0.0507
RMSE	Mean:	0.0819	0.0821	0.1872	0.2330	0.2676
	St. Dev.:	0.0752	0.0736	0.1631	0.2507	0.2458
	Min:	0.0173	0.0173	0.0332	0.0258	0.0232
	Max:	0.2697	0.2806	0.6193	1.1064	1.0827
MAE	Mean:	0.0545	0.0554	0.1111	0.1143	0.1893
	St. Dev.:	0.0528	0.0507	0.1019	0.1110	0.1581
	Min:	0.0100	0.0105	0.0201	0.0207	0.0138
	Max:	0.2012	0.2003	0.4050	0.4537	0.6402

* In Table 3, the error for each country-month is defined as the forecast value minus the true value of the log black market exchange rate. The country-specific variable in the left-hand column is reported for each model, with the mean across the sample of countries reported with its standard deviation, minimum, and maximum values computed across all 34 countries. For model (5), Egypt is excluded from the sample and Jamaica is included but with 43 versus 48 data points, since model (5) needed 5 additional months to converge in the case of Jamaica.

Table 4: Sample performance statistics for out-of-sample fit and forecasting exercises, models (1) – (5).

	Model:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	Number of Countries	34	34	34	34	33
Panel A. "Ex-post" Out-of-sample Fit (Ex-post values of independent variables known)						
Cumulative Monthly Return (APR)	Mean:	-0.055	0.048	0.311	0.357	0.176
	St. Dev.:	0.327	0.258	0.437	0.446	0.209
	Min:	-1.636	-0.235	-0.006	0.098	0.002
	Max:	0.297	1.316	2.149	2.076	0.985
Theil U-statistic results (baseline of Random Walk)	RW outperforms:	-	0	24	14	29
	Insignificant diff.	-	33	7	14	4
	Alt. model outperforms	-	1	3	6	0
Theil U-statistic results (baseline of Random Walk with Drift)	RWWD outperforms:	1	-	25	15	30
	Insignificant difference	33	-	6	13	3
	Alt. model outperforms	0	-	3	6	0
Proportion of Correct Direction Forecasted	Mean:	0.576	0.566	0.667	0.701	0.608
	St. Dev.:	0.150	0.134	0.100	0.073	0.088
	Min:	0.167	0.342	0.472	0.563	0.477
	Max:	0.889	0.889	0.894	0.851	0.875
Anatolyev-Gerko Test	EP>0, significant at 5% level	13	2	16	12	11
	EP insignificantly different from zero	1	5	7	9	12
	EP<0, significant at 5% level	20	27	11	13	10
Panel B. "Ex-ante" Out-of-Sample Forecasting (Ex-post values of independent variables unknown)						
Cumulative Monthly Return (APR)	Mean:	-0.055	0.048	0.003	-0.020	0.170
	St. Dev.:	0.327	0.258	0.193	0.178	0.249
	Min:	-1.636	-0.235	-0.213	-0.869	-0.045
	Max:	0.297	1.316	0.956	0.241	1.028
Theil U-statistic results (baseline of Random Walk)	RW outperforms:	-	0	33	33	29
	Insignificant diff.	-	33	1	1	3
	Alt. model outperforms	-	1	0	0	1
Theil U-statistic results (baseline of Random Walk with Drift)	RWWD outperforms:	1	-	33	33	29
	Insignificant difference	33	-	1	1	3
	Alt. model outperforms	0	-	0	0	1
Proportion of Correct Direction Forecasted	Mean:	0.576	0.566	0.541	0.534	0.588
	St. Dev.:	0.150	0.134	0.082	0.078	0.085
	Min:	0.167	0.342	0.333	0.333	0.417
	Max:	0.889	0.889	0.750	0.722	0.813
Anatolyev-Gerko Test	EP>0, significant at 5% level	13	2	2	2	13
	EP insignificantly different from zero	1	5	15	17	11
	EP<0, significant at 5% level	20	27	17	15	9

*In Table 4, the error for each country-month is defined as the forecast value minus the true value of the log black market exchange rate. The country-specific variable in the left-hand column is reported for each model, with the mean across the sample of countries reported with its standard deviation, minimum, and maximum values computed across all 34 countries, except in the case of column (5). For model (5), Egypt is excluded from the sample and Jamaica is included but with 43 versus 48 data points, since model (5) needed 5 additional months to converge in the case of Jamaica. There were the following cases of bankruptcy based on realized trading strategy returns: for Panel A, model (1), Morocco registered bankruptcy in month 82, and Myanmar registered bankruptcy in month 15; for model (2), Morocco registered bankruptcy in month 82, and Myanmar in month 15; for model (5), Morocco registered bankruptcy in month 82. For panel B, model (1) Morocco registered bankruptcy in month 82, and Myanmar in month 15; for model (2), Morocco registered bankruptcy in month 82, and Myanmar in month 15; for model (4), Myanmar registered bankruptcy in month 15; for model (5), Myanmar registered bankruptcy in month 15.