Is Arbitrage Tying the Price of Ethanol to that of Gasoline? Evidence from the Uptake of Flexible-Fuel Technology

ONLINE APPENDIX:
FURTHER DETAILS ON DATA AND CAUSALITY TESTS

Alberto Salvo
Northwestern University (KSM)
Cristian Huse
Stockholm School of Economics

October 2010

1 Data

We detail the sources of data used in the paper, as well as provide further descriptive statistics.

Domestic fuel prices A monthly state-level panel of retail prices for gasoline and ethanol, in current local currency R$/liter, over the period Jul-2001 through Sep-2009, covering all 27 states of the Brazilian federation, was obtained from the National Agency for Oil (ANP). For each month-state pair, ANP reports the mean per-liter price paid by the consumer (i.e., including sales taxes) across a representative sample of retail fueling stations in that state. (In sharp contrast to the U.S., virtually all stations carry both gasoline and ethanol.)

Current prices were then converted to constant R$/liter using the Brazilian Institute for Geography and Statistics' (IBGE) monthly national Consumer Price Index (the "IPCA").

To complement Figure 6 in the paper (summarizing how the distribution of relative ethanol prices, p^e/p^g, across the 27 states has evolved over time), Figure A1 provides information on the cross-sectional distribution of prices for the *individual* fuels: gasoline, p^g, in the left panel and ethanol, p^e, in the right panel. For every month, the figure indicates the maximum, the 75th percentile, the 25th percentile, and the minimum of the price distribution across the 27 states. For example, in July 2001 (the first month of the sample), the highest gasoline price was 3.16 R$/l (the state of Pará), the lowest was 2.77 R$/l (the state of Goiás) and the interquartile range (marked with the thick lines) was 2.92 - 2.83 R$/l; the highest ethanol price was 2.26 R$/l (the state of Acre), the lowest was 1.50 R$/l (the state of São Paulo) and the interquartile range was 2.07 - 1.76 R$/l. As mentioned in the paper, notice that the spatial price dispersion in gasoline (about 0.60 R$/l from max to min) has consistently been markedly lower than in ethanol (about 1.00 R$/l). This probably reflects the presence of non-market forces in gasoline, namely the government's determination to subsidize gasoline in the more remote and less-developed (northern/northeastern) states by not fully passing through the cost of transporting gasoline from oil refineries, which are predominantly located along the coastline (mainly in the richer southeast and south of the country, but also in some northeastern states such as Bahia and Ceará). In setting (wholesale) gasoline prices the government seems to adopt a strategy of "uniform pricing" to some degree, particularly in comparison to the spatial variation we observe in ethanol prices across the different regions of the country.

World oil and sugar prices Monthly "world" oil prices (spot prices of West Texas Intermediate crude oil at Cushing, OK, in current US$/barrel), over the period Jul-2001 through Sep-2009, were obtained from the Energy Information Administration (U.S. Department of Energy).

Monthly world sugar prices (International Sugar Agreement prices, in US$ cents/pound), over the same period, were obtained from the International Sugar Organization (ISO).[1]

Since we wish to make world oil and sugar prices comparable with domestic fuel prices, we follow two alternative approaches: (i) convert all series into constant US$, and (ii) convert all series into constant R$. In approach (i) (i.e., constant US$), we begin by converting current domestic prices in R$ into current US$ using current R$/US$ exchange rates (monthly daily averages obtained from the Central Bank of Brazil, BACEN), and then convert both current domestic prices in US$ and current world prices in US$ into constant US$ using the Bureau of Labor Statistics' (U.S. Department of Labor) Consumer Price Index (averaged across U.S. cities and all items). In approach (ii) (i.e., constant R$), we begin by converting current world prices in US$ into current R$ using current R$/US$ exchange rates, and then convert both current domestic prices in R$ and current world prices in R$ into constant R$ using IBGE's CPI (see above).

[1] Alternatively, a quarterly export price index for Brazilian sugar, in current US$ terms, compiled by the Center for the Study of Trade (FUNCEX), was obtained from Brazil's Institute for Applied Economic Research (IPEA). It is reassuring that (accounting for the different levels of time series aggregation) the FUNCEX export price index and the ISO price series are highly collinear.

Other data used to test assumptions and robustness, or to prepare figures
A monthly nationwide wholesale price index for refined sugar, in current R$ terms, was available from the Fundação Getulio Vargas (FGV) over the period Jul-2001 to Dec-2008, which we then complemented with a similar but more recent time series through Sep-2009 obtained from the Center for Advanced Studies in Applied Economics (CEPEA/ESALQ).

Monthly producer prices for hydrated ethanol in the state of São Paulo, in current R$/liter (FOB ethanol mill and net of sales taxes), over the period Jul-2001 to Sep-2009, were obtained from the Brazilian Sugarcane Industry Association (UNICA).

Weekly producer prices for (unblended, or pure "type A") gasoline in the southeastern region of Brazil (which includes the state of São Paulo), in current R$/liter (an average across refineries and importers and net of ICMS sales tax), over the period Jan-2002 to Sep-2009, were made available by ANP. To compute a monthly series from the weekly producer gasoline prices we took the mean of prices in the first and last weeks of each month. Further, since producer gasoline prices were available only for Jan-2002 onward, for the 6 months between Jul-2001 and Dec-2001 we extrapolated backwards using the monthly variation in the mean purchase price for gasoline paid by retail fueling stations in the state of São Paulo, which was available from ANP for those 6 months.

All current prices (or current price indices) were converted to constant R$/liter (or constant R$ terms) as explained above.

State-level GDP for the three most recently available years (2005 to 2007) was obtained from IBGE.

The tonnage of sugar cane harvested by state over three harvesting cycles (2005/06 to 2007/08) was obtained from UNICA.

To construct a proxy for state-level FFV penetration, we obtained the number of passenger cars in each state's (Department of Motor Vehicles') official registry, and how this has varied since 2001, from the National Department for Transport (DENATRAN, who acts as an information aggregator). Different sources (e.g., the Brazilian Automobile Dealers Association, Fenabrave, and the Brazilian Autoparts Industry Association, Sindipeças) had warned us that Denatran's fleet sizes are inflated, particularly for earlier vintages (i.e., scrappage rates are unrealistically low). To Denatran's fleet we applied Sindipeças' estimated scrappage rates by car age (around 1.8% per year in operation), which their "Circulating Fleet Study Group" infers from car parts markets and insurance markets. Since Denatran's fleet figures are not available by fuel type, we used Sindipeças' data on the proportion of cars added to the nationwide fleet every year that were FFVs, which Sindipeças bases on new car sales (i.e., we assume similar FFV shares of new car sales across states, though the number of new car sales varies widely). Further details can be provided upon request.

2 Causality tests

We start investigating the time series properties of the (constant R$/liter) price series by studying their order of integration, fuel-by-fuel and state-by-state. To allow for "structural breaks" in the series, we use the unit root test proposed by Saikkonen and Lütkepohl (2002) and Lanne et al (2002). Throughout our analysis, we select the optimal number of lags according to the Schwarz criterion, and allow for seasonal components (monthly dummies). It is worth pointing out that the number of time periods—i.e., 99 months between July 2001 and September 2009—is low relative to that wished for by a time series econometrician, with non-trivial implications on testing power, but data availability and government control of ethanol prices until 1999 restrict us from using a longer time series.

For a given state-fuel price series, if the null hypothesis of non-stationarity of the unit root test is rejected, we conclude that the series is I(0). If the series is not found to be stationary in levels, we analyze its first difference; for all such series where we performed the unit root test on first differences, we rejected the null of non-stationarity, thus obtaining that the series in levels is I(1).

For those states where both fuel price series are I(0), we run a VAR in price levels and test for both instantaneous and Granger causality. For those states where both fuel price series are I(1), we test for cointegration using the Johansen (1991) test. For the case where the two I(1) price series cointegrate, we estimate a Vector Error Correction Model and test for instantaneous and Granger causality using the corresponding VAR representation. For the case where the two I(1) price series do not cointegrate, we test for causality by estimating a VAR in the first difference of the price series.

A time series Z is said to Granger-cause another time series Y if it can be shown that lagged values of Z provide statistically significant information about future values of Y. In our setting, this corresponds to estimating (assuming stationary price series)

$$\begin{bmatrix} p_t^e \\ p_t^g \end{bmatrix} = \sum_{k=1}^{K} \begin{bmatrix} \alpha_{11,k} & \alpha_{12,k} \\ \alpha_{21,k} & \alpha_{22,k} \end{bmatrix} \begin{bmatrix} p_{t-k}^e \\ p_{t-k}^g \end{bmatrix} + X_t + \begin{bmatrix} u_t^e \\ u_t^g \end{bmatrix}$$

where K is determined using the Schwarz criterion and X_t corresponds to variables that are exogenous at time t (namely, a level shift whenever this was significant in the separate Saikkonen-Lütkepohl unit root test, as well as a full set of monthly dummies). In the above specification, p^g does not Granger-cause p^e if, and only if, $\alpha_{12,k} = 0, k = 1, 2, ..., K$ (this is the null hypothesis). Similarly, p^e does not Granger-cause p^g if, and only if, $\alpha_{21,k} = 0, k = 1, 2, ..., K$. Intuitively, if the coefficients on lagged variables $\alpha_{21,k}$ are jointly non-significant, then p^e does not have explanatory power for future values of p^g. For example, recalling Table 2, in the case of 5 states (ES, RN, SE, DF and RR) we find that parameters $\alpha_{12,k}$ tend to be significantly different from zero (i.e., evidence of

p^g Granger-causing p^e) whereas parameters $\alpha_{21,i}$ tend to be jointly non-significant (i.e. no evidence of p^e Granger-causing p^g).

Instantaneous causality is characterized by non-zero correlation of the error terms u^e and u^g. The null hypothesis $E(u_t^e, u_t^g) = 0$ is tested against the alternative of non-zero covariance (see Lütkepohl 1991 for a discussion of causality tests). Again recalling the results of Table 2, we find overwhelming support for instantaneous causality between p^e and p^g, indicating the contemporaneous correlation between shocks impacting the two price series.

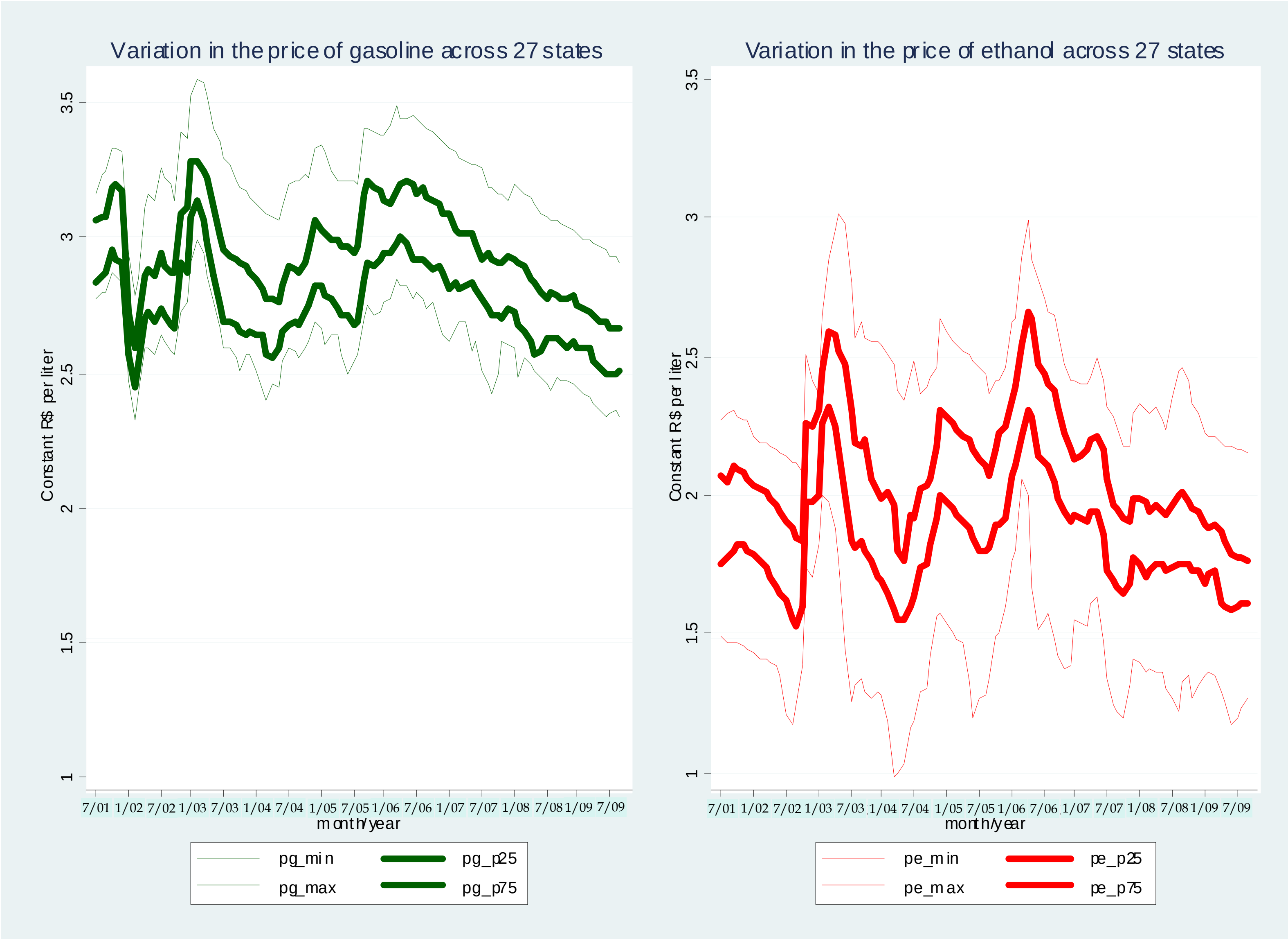

Figure A1: Evolution of the cross-sectional variation (i) in the price of gasoline (left panel), and (ii) in the price of ethanol (right panel), at the pump. For each month in the period Jul-2001 to Sep-2009, the maximum, the minimum, and the interquartile range of prices across 27 states are shown. Prices in constant R$ per liter (Brazil CPI base Sep-2009). Source: ANP and IBGE.