Center for Quantitative Economics

Measurement Errors in Index Trader Positions Data: Is the Price Pressure Hypothesis Still Invalid?

Martin T. Bohl [†], Nicole Branger [†] und Mark Trede [†]

80/2019

[†] Department of Economics, University of Münster, Germany

wissen.leben
WWU Münster

Measurement Errors in Index Trader Positions Data: Is the Price Pressure Hypothesis Still Invalid?*

Martin T. Bohl† Nicole Branger‡ Mark Trede§

March 18, 2019

Abstract

In this paper, we examine whether the repeated rejection of Masters' price pressure hypothesis is robust with respect to measurement errors in index trader position data. We allow for autocorrelated errors and a potential impact of index trader positions on the level and volatility of commodity returns. The resulting state-space model is estimated via particle MCMC. The empirical investigation relies on weekly data for eleven commodities contained in the SCoT reports. Our empirical findings show that the rejection of the price pressure hypothesis is robust concerning the inclusion of measurement errors in index trader positions data.

JEL Classification: C18, G41, Q02, Q14

Keywords: Masters' Price Pressure Hypothesis, Measurement Errors, Commodity Futures Markets, Index Traders, CFTC Data

*Acknowledgements: The authors would like to thank Alexander Pütz, Martin Stefan, Christoph Sulewski and Claudia Wellenreuther for helpful comments.

†Westfälische Wilhelms-University Münster, Department of Economics, Am Stadtgraben 9, 48143 Münster, Germany, martin.bohl@wiwi.uni-muenster.

‡Westfälische Wilhelms-University Münster, Finance Center Münster, Universitätsstraße 14–16, 48143 Münster, Germany, nicole.branger@wiwi.uni-muenster.de.

§Corresponding author. Westfälische Wilhelms-University Münster, Department of Economics, Institute of Econometrics, Am Stadtgraben 9, 48143 Münster, Germany, mark.trede@uni-muenster.de. Tel: +49-251-8325006.

1 Introduction

Masters' (2008) well-known price pressure hypothesis blames commodity index traders' investments for driving up commodity futures prices throughout the 2000s. Due to its intuitively appealing argumentation the Masters' hypothesis has received broad public dissemination via mass media and the NGO scene. With a few exceptions (Gilbert (2010), Tang and Xiong (2012), Singleton (2011)), numerous empirical studies overwhelmingly reject the hypothesis of an influence of index traders' investments on commodity price dynamics (Stoll and Whaley (2010), Aulerich, Irwin and Garcia (2012), Irwin and Sanders (2012), Hamilton and Wu (2015), Sanders and Irwin (2011), among others).

While the empirical evidence against the price pressure hypothesis is fairly robust concerning different data sets, econometric techniques, sample periods and model specifications, one objection against the aforementioned empirical investigations can be made: The empirical analyses rely on position data of index traders sourced from the U.S. Commodity Futures Trading Commission's (CFTC) commitments of traders (CoT) data base. It is well-known that the CFTC's position data are infected with measurement errors. Concerning the traditional CoT data a frequent complaint is that speculators may have an incentive to self-classify their activity as commercial hedging to circumvent speculative position limits. In contrast, little incentives exist for traders to be classified as non-commercials. Moreover, in the non-reporting category the composition of commercial and non-commercial traders is unknown (e.g., Ederington and Lee (2002), Sanders, Boris and Manfredo (2004)).

As a response to the rapid increase of long-only index traders' investments the CFTC released Supplemental Commitments of Traders (SCoT) reports which contain long- and short-positions held by index traders for twelve agricultural markets. Data are publicly available on a weekly basis since January 3, 2006. However, the SCoT classification also has deficits concerning index traders' positions. The CFTC states that "...some traders assigned to the index traders' category are engaged in other futures activity that could not be disaggregated. ...Likewise, the index traders' category will not include some traders who are engaged in index trading, but for whom it does not represent a substantial part of their overall trading activity" (CFTC (2008)).

Another source of measurement errors in index traders' position data arises from the technique of commodity index investing because most investments undertaken by swap dealers use the over-the-counter market. Swap dealers offset their risk position in the futures market and commonly net their customers' long and short positions before placing hedges in the futures market. Due to the internal netting of underlying positions in the swap market only a fraction of positions appears in the futures market, and, hence, in the SCoT reports (Stoll and Whaley (2010), Sanders, Irwin and Merrin (2010)).

Certainly, the SCoT data are an improvement over the heavily aggregated traditional CoT classification and provide new information about the trading activity of long-only index traders. Nevertheless, the above arguments indicate that index fund positions are still contaminated by measurement errors which may have an influence on the empirical findings of the price pressure hypothesis, in particular, its overwhelming rejection. The main contribution of this paper is an empirical analysis of the impact of measurement errors in index traders' positions on the evidence against the price pressure hypothesis. To the best of our knowledge, measurement errors have not been systematically taken into account in regressions designed to study the influence of index traders' investments on commodity returns so far. The question is thus whether the rejection of the price pressure hypothesis is due to data limitations or due to a truly non-existent price impact of index fund positions.

Our empirical investigation relies on weekly futures prices and position data from January 3, 2006 to May 29, 2018 for eleven commodities covered in the SCoT reports. We assume that observed relative positions are noisy versions of the true relative positions. Based on a model which allows for autocorrelated measurement errors, we test whether index traders' positions have an impact on returns and volatilities of commodity futures. The parameters are estimated via MCMC. For none of the commodities, we find a significantly non-zero impact of the true unobservable index trader positions on commodity returns and return volatilities. Overall, our results confirm the rejection of the price pressure hypothesis found in the aforementioned studies.

The remainder of this paper is organized as follows. In Section 2, we outline the econometric method. Section 3 discusses the data and the empirical results. Finally, Section 4 summarizes and concludes.

2 Econometric method

Classical measurement errors have always been an integral part of econometric models. However, the independence assumption for classical measurement errors is often unrealistic. An overview of recent advances in dealing with non-classical measurement errors can be found in Schennach (2016) or the special issue of the Journal of Econometrics (see the introduction by Hu and Wansbeek, 2017). Often, explicitly modelling some non-classical properties of the errors allows identification. This is also the route that we will pursue.

We assume that the index traders position is subject to measurement errors. If the real proportion of index traders is higher than observed, it is very likely that it will also be higher in the following period. Such an autoregressive structure for the measurement error helps to identify and estimate the model parameters.

We consider a linear mean equation for returns

$$r_t = \alpha + \sum_{p=1}^{P} \beta_p r_{t-p} + \sum_{q=0}^{Q} \gamma_q \tilde{N}_{t-q} + v_t \tag{1}$$

where $r_t = \ln P_t - \ln P_{t-1}$ is the log-return, P_t the futures price and $\tilde{N}_t$ the true, but unobserved, net long position of index traders. The net long position is subject to measurement error u_t. The measured net long position is

$$N_t = \tilde{N}_t + u_t.$$

We assume that the latent measurement error u_t follows an autoregressive process,

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{2}$$

with constant $Var(\varepsilon_t) = \sigma_\varepsilon^2$. The zero-mean assumption for the errors is not restrictive, since any systematic bias in the measured net long position could be merged into the intercept α in equation (1).

In a state-space representation, the measurement error is a state variable and (2) is its transition equation. The observation equations are (1) and

$$\tilde{N}_t = \begin{cases} 0 & \text{if } N_t - u_t < 0 \\ N_t - u_t & \text{if } 0 \leq N_t - u_t \leq 1 \\ 1 & \text{if } N_t - u_t > 1 \end{cases} \tag{3}$$

The proportion of index traders' net long positions cannot lie outside the unit interval. We impose this restriction even though it is hardly ever binding in our empirical application.

The proportion of index traders' net long positions might also affect volatility. We assume the stochastic volatility specification

$$v_t \sim N(0, h_t)$$

with

$$(\ln h_t - \omega) = \phi\,(\ln h_{t-1} - \omega) + \sum_{k=0}^{K} \lambda_k \tilde{N}_{t-k} + \eta_t \tag{4}$$

and $\eta_t \sim N(0, \sigma_\eta^2)$. The volatility h_t is the second state variable, and (4) is its transition equation.

The null hypothesis that the net long positions neither have an impact on the level nor on the volatility of the returns can be formalized as

$$H_0 : \gamma_q = 0 \text{ for } q = 0, \ldots, Q \text{ and } \lambda_k = 0 \text{ for } k = 0, \ldots, K. \tag{5}$$

Due to the potential impact of $\tilde{N}$ on volatility, the model is nonlinear and, hence, the classical Kalman filter is not applicable. We suggest to estimate the model parameters

$$\theta = (\alpha, \beta_1, \ldots, \beta_P, \gamma_0, \ldots, \gamma_Q, \omega, \phi, \lambda_0, \ldots, \lambda_K, \rho, \sigma_\varepsilon, \sigma_\eta)$$

by particle MCMC as suggested in Flury and Shephard (2011). We apply the DEMC algorithm (Ter Braak, 2006) to generate proposals for Metropolis-Hastings steps. It is designed for parallel computing and substantially speeds up drawing from the posterior distributions. The larger the number of parallel chains, the faster convergence to the posterior distribution. Deleting a burn-in phase for each chain, the draws of all chains can be merged. They represent the posterior distribution of the parameter vector θ.

In our empirical application, we set the number of lags to $K = 0$. Drop some subindices we arrive at a simplified model consisting of (3), (2) and

$$
\begin{aligned}
r_t &= \alpha + \beta r_{t-1} + \gamma \tilde{N}_t + v_t \\
\ln h_t &= \omega + \phi(\ln h_{t-1} - \omega) + \lambda \tilde{N}_t + \eta_t
\end{aligned}
$$

and $v_t \sim N(0, h_t)$, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $\eta_t \sim N(0, \sigma_\eta^2)$. The parameter vector is

$$\theta = (\alpha, \beta, \gamma, \omega, \phi, \lambda, \rho, \sigma_\varepsilon, \sigma_\eta).$$

The prior distributions are typically set up as independent, and often as uninformative. Since the measured, as well as the latent true net long position are defined as proportions, they must lie within $[0, 1]$. We impose the following rather uninformative prior restrictions (for weekly data)

$$
\begin{array}{lll}
|\alpha| < 0.01 & |\beta| < 0.5 & |\gamma| < 0.1 \\
|\omega| < 20 & \phi \in [0, 1] & |\lambda| < 0.1 \\
|\rho| < 0.9 & \sigma_\varepsilon \in [0, 1] & \sigma_\eta \in [0, 1].
\end{array}
$$

3 Data and empirical findings

We proceed to investigate if the Masters' hypothesis is still invalid when a measurement error is explicitly taken into consideration. To do so, we estimate the model parameters θ and test the null hypothesis (5) with $K = 0$ for eleven commodities: cocoa, coffee, corn, cotton, feeder cattle, lean hogs, live cattle, soybean oil, soybeans, sugar and wheat-srw. Data on futures prices are collected from Thomson Reuters Datastream, where continuous futures price time series are constructed by switching the contract on the first trading day of the expiring month. To match the position data Tuesday-to-Tuesday logarithmic price differences are calculated.

The commodity index traders (CIT) long and short positions are provided by CFTC.[1] All time series range from 3 January 2006 to 29 May 2018.

Table 1 reports descriptive statistics of the weekly returns and the net long positions. About half of the commodities have a negative skewness. The kurtosis is larger than 3 for all commodities

[1]See http://www.cftc.gov/MarketReports/CommitmentsofTraders/HistoricalCompressed/index.htm.

indicating the presence of heavy tails and outliers. For most commodities, the net long position oscillates roughly between 50 and 100 percent with no clear time trend. The smallest proportions are observed for cocoa and sugar. The upper bound of 100 percent is reached by seven series but figure 1 reveals that hitting the 100 percent bar is an exception.

	weekly returns				net long position		
	mean	stddev	skew	kurtosis	mean	min	max
Cocoa	0.08	3.91	0.01	4.10	73.8	14.1	100.0
Coffee	0.02	4.27	0.13	3.72	83.2	49.8	100.0
Corn	0.10	4.35	-0.11	5.34	78.3	50.8	98.4
Cotton	0.08	4.40	-0.84	8.86	86.5	47.2	100.0
Feeder cattle	0.04	2.41	-0.26	5.34	85.0	56.4	100.0
Lean hogs	0.02	4.61	-0.30	6.73	90.7	69.2	100.0
Live cattle	0.01	2.48	-0.53	6.59	94.4	81.7	100.0
Soybean oil	0.06	3.36	0.09	4.15	83.0	58.6	100.0
Soybeans	0.08	3.52	-0.49	4.44	76.1	41.0	99.1
Sugar	-0.02	4.82	0.09	4.35	73.8	29.7	95.4
Wheat	0.07	4.76	0.34	4.14	75.9	56.3	99.0

Table 1: Descriptive statistics of weekly returns and net long positions (all in percent)

We estimate the posterior distribution of θ for all commodities using the DEMC algorithm. The number of parallel Markov chains is 64. The number of iterations is 30000. We delete the first 10000 draws from each chain as burn-in phase and thin the remaining 20000 draws keeping only every 20th value to reduce dependence. Hence, the total number of draws for each parameter is $64 \times 20000/20 = 64000$. Figure 2 shows the posterior distributions of all model parameters and all commodities as violin plots. The small horizontal lines are the posteriors' 0.05 and 0.95 quantiles.

The plot reveals that for many parameters all credibility intervals cover zero. Concerning the dynamics of returns, this is true for α, γ, and λ. The β intervals cover zero for all commodities save corn (negative) and lean hogs (positive). In general, the mean equations reduce to $r_t = v_t$. The net long position does not have any discernible influence on returns.

The parameters of the latent measurement error process, ρ and σ_ε, are only poorly identified. Their posteriors are close to the priors.

As to volatility, only the usual stochastic volatility parameters ω, ϕ and σ_η are significantly different from zero. The generally large values of ϕ show the persistence of volatility clusters. The posteriors of λ show no impact of the net long position on volatility. The only commodity

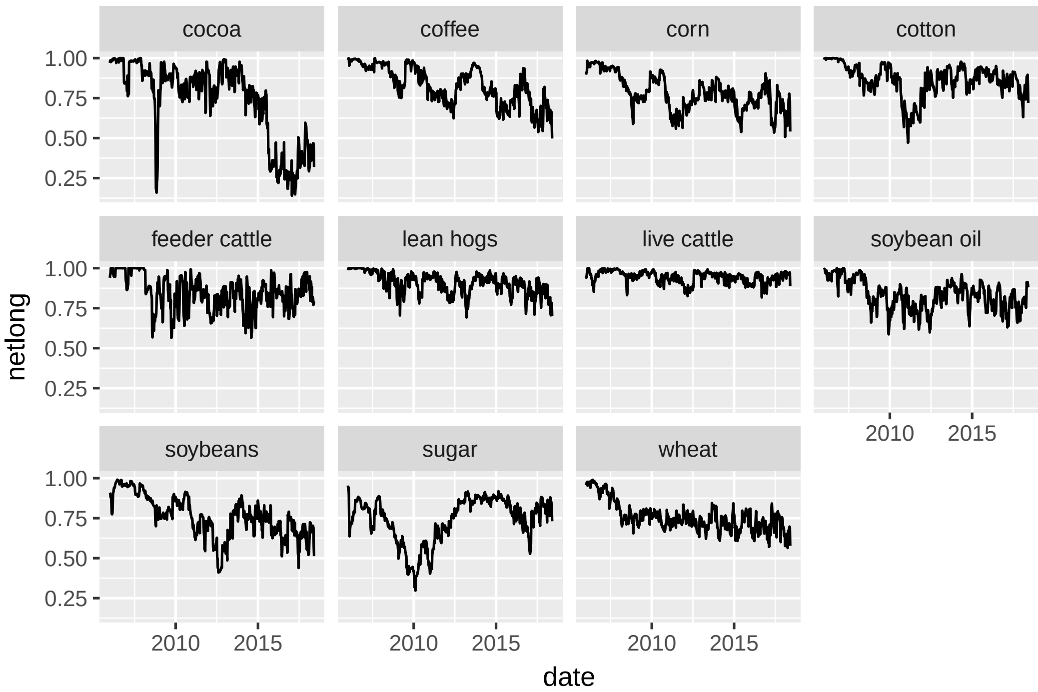

Figure 1: Weekly net long position of index traders in the commodities

where the distribution of λ shifts away from zero (although not significantly) is sugar. Here, a larger net long position tends to decrease volatility.

Hence, neither can we find any impact of the net long position on the expected return, nor on the return volatility. Taking measurement error into account does not change the invalidity of the Masters' hypothesis.

4 Summary and conclusions

Masters' (2008) price pressure hypothesis claims that the run-up in commodity prices during the 2000s was driven to a substantial degree by capital inflows of commodity index investments into futures markets. However, the majority of empirical studies find evidence against this hypothesis. Most of these investigations rely on data contained in the CFTC's SCoT reports. The position data for index traders are subject to a number of limitations which may limit the degree of confidence one can place in the empirical findings against Masters' price pressure hypothesis.

In this paper, we take into account measurement errors in index traders' position data as a potential reason for the repeated rejection of the price pressure hypothesis. The empirical investigation exploits weekly data on futures prices and index trader positions over the period from January 3, 2006 to May 29, 2018 for eleven commodities contained in the SCoT reports. Our empirical findings show that the rejection of Masters price pressure hypothesis is robust concerning the inclusion of measurement errors in index trader position data. Our empirical findings strengthen the results of studies relying on SCoT's index trader data.

References

[1] Aulerich, N.M., Irwin, S.H. and Garcia, P. (2012). Bubbles, Food Prices, and Speculation: Evidence from the CFTC's Daily Large Trader Data Files. Working Paper.

[2] Commodity Futures Trading Commission (2008). Staff Report on Commodity Swap Dealers and Index Traders with Commission Recommendations.

[3] Ederington, L. and Lee, J.H. (2002). Who Trades Futures and How: Evidence from the Heating Oil Market. *Journal of Business* 75, 353-373.

[4] Flury, T., Shephard, N. (2011). Bayesian Inference Based only on Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models. *Econometric Theory* 27, 933-956.

[5] Gilbert, C.L. (2010). Speculative Influences on Commodity Futures Prices 20062008. Working Paper, University of Trento.

[6] Hamilton, J.D. and Wu, J.C. (2015). Effects of Index-Fund Investing on Commodity Futures Prices. *International Economic Review* 56, 187-205.

[7] Hu, Y., Wansbeek, T. (2017). Measurement error models: Editors' introduction. *Journal of Econometrics* 200, 151-153.

[8] Irwin, S.H. and Sanders, D.R. (2012). Testing the Masters Hypothesis in Commodity Futures Markets. *Energy Economics* 34, 256-269.

[9] Masters, M.W. (2008). Testimony before the Committee on Homeland Security and Governmental Affairs, United States Senat.

[10] Sanders, D.R., Boris, K. and Manfredo, M. (2004). Hedgers, Funds, and Small Speculators in the Energy Futures Markets: An Analysis of the CFTC's Commitments of Traders Reports. *Energy Economics* 26, 425-445.

[11] Sanders, D.R. and Irwin, S.H. (2011). New Evidence on the Impact of Index Funds in US Grain Futures Markets. *Canadian Journal of Agricultural Economics* 59, 519-532.

[12] Sanders, D.R., Irwin, S.H. and Merrin, R.P. (2010). The Adequacy of Speculation in Agricultural Futures Markets: Too Much of a Good Thing? *Applied Economics Perspectives and Policy* 32, 77-94.

[13] Schennach, S.M. (2016). Recent Advances in the Measurement Error Literature. *Annual Review of Economics* 8, 341-377.

[14] Singleton, K.J. (2011). Investor Flows and the 2008 Boom/Bust in Oil Prices. Working Paper Stanford University.

[15] Stoll, H.R. and Whaley, R.E. (2010). Commodity Index Investing and Commodity Futures Prices. *Journal of Applied Finance* 20, 7-46.

[16] Tang, K. and Xiong, W. (2012). Index Investment and Financialization of Commodities. *Financial Analysts Journal* 68, 54-74.

[17] Ter Braak, C.J.F. (2006). A Markov Chain Monte Carlo Version of the Genetic Algorithm Differential Evolution: Easy Bayesian Computing for Real Parameter Spaces. *Statistics and Computing* 16, 239-249.

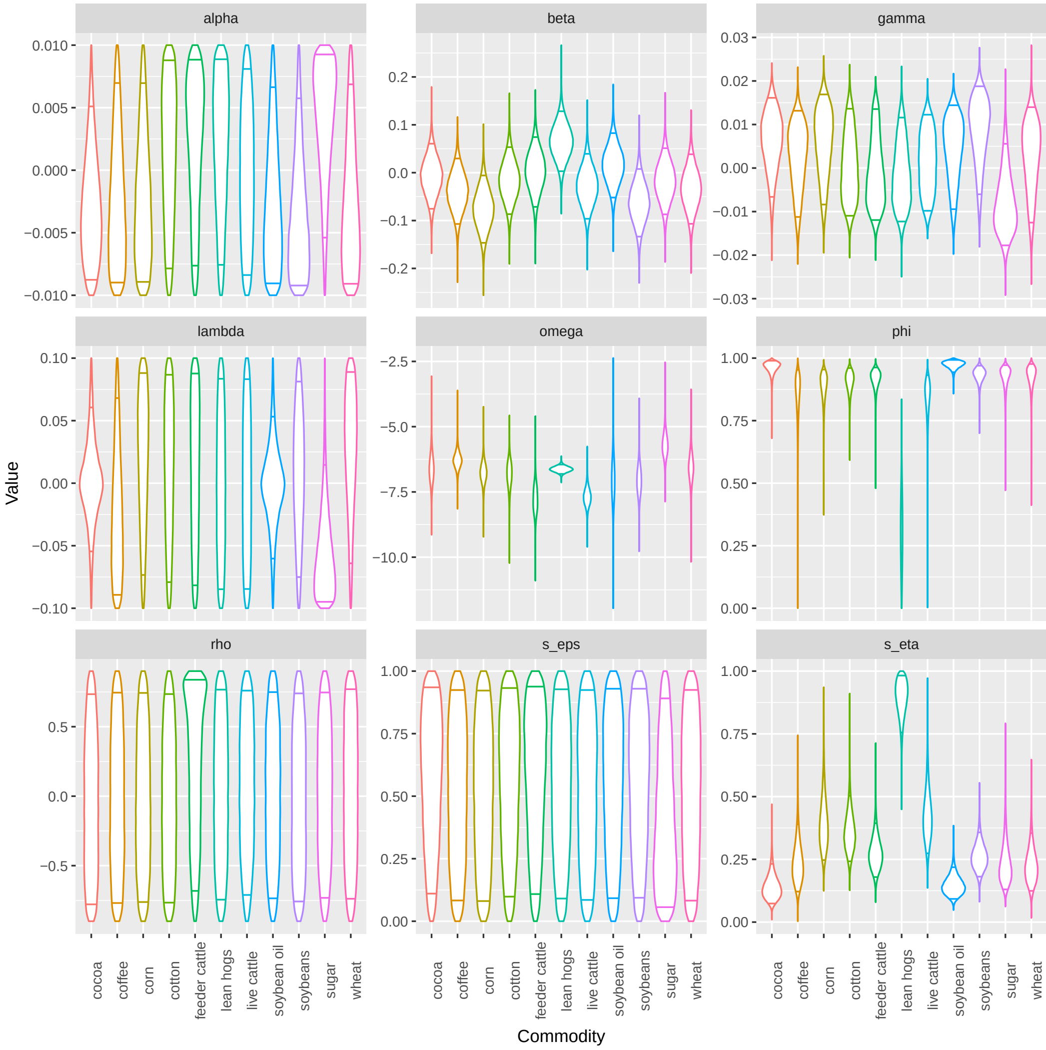

Figure 2: Posterior densities of all model parameters for all commodities