“Negligible” Trends, Spurious Linearity

Juan R. A. Bobenrieth, Eugenio S. A. Bobenrieth, Brian D. Wright and Di Zeng

June 23, 2017

Abstract

Neglect of trends has unrecognized implications for interpretation of autoregressions. In nonparametric analysis of instructive samples of commodity price series, correspondences relating current price to the following price are linear, and returns appear independent of current price. After detrending current price, these relationships appear highly nonlinear, suggestive of the implications of a model of storage arbitrage. In such a model, any trend is not revealed in expected returns on positive stocks but in expected jumps from boom prices. We implement a new approach to consistent estimation of nonlinear empirical models with a trend in price that might not be zero.
When primary commodity prices are high, exporting countries confront a strategic choice. Should they make multiyear commitments to expanded commodity production, or divert the windfall profits to alternative, consumption-smoothing investments? The answers to questions like this depend on the nature of the relation between current price and price realizations over the life of the project.

During the commodity boom after the end of World War Two, Singer (1950) and Prebisch (1950) echoed the warning of Kindleberger (1943, p.349) that post-war investments in primary commodity production might be of “fleeting profitability” and that commodity prices exhibit a secular decline in their terms of trade. They urged commodity exporting developing countries to invest profits from a post-war boom in industrialization rather than in expansion of primary commodity production (Singer 1950, p.477).

Their advice had a controversial influence on postwar industrialization strategies, but it also initiated a large empirical literature on the “Prebisch-Singer hypothesis” of a secular decline in the relative prices of primary products. Those papers that confirmed the hypothesis generally found trends so small as to appear to be economically insignificant. As summarized by Cashin and McDermott (2002, p.175), “Although there is a downward trend in real commodity prices, this is of little practical policy relevance, since it is small and completely dominated by the variability of prices.” Neglect of trends in economic analysis of the policies related to the shorter-term behavior of commodity prices has seemed empirically justified.

In this paper we contest this conclusion, showing that neglected small trends can linearize price autoregressions, inducing the incorrect inference that returns are independent of current price.

We first show, via nonparametric exploration of instructive samples of prices, that a neglected long run annual trend that constitutes a minor part of typical year-to-year price variation can dominate inferences about persistence and volatility of prices. It can make returns and their volatility appear to be constant functions of price, and exaggerate persistence of price booms.

Consider Figure 1, which relates the logarithms of current real annual prices to the immediately following observations for samples of cotton and maize prices. We choose a sample interval, 1960-2007, for reasons discussed below.

For both cotton and maize, the correspondence between current log price and the subsequent log price realization shown in Figure 1 appears strictly increasing and approximately linear, with high first order sample correlations of prices, 0.93 and 0.90 respectively. Further, prices appear centered close to the forty-five degree line, supporting the conclusion of Cashin and McDermott that any trend appears to account for only a negligible of short-run price movements.

Given high positive autocorrelation and negligible trends, this linear characterization supports two inferences of Deaton (2010 pp. 7-10): first, and contrary to the advice of Kindelberger (1943), saving of excess export returns during commodity price booms appears no more justified than borrowing during price slumps. Second, the principal economic model of commodity storage arbitrage with i.i.d. harvest disturbances, originated
Figure 1: Sample real prices and their immediately following real prices (in natural logarithms), deflated using the Manufactures Unit Value Index, (1960-2007)
by Gustafson (1958), fails a “transparent test” of one key implication: the autoregression should “flatten out” at high prices. This test is very attractive, in that it completely bypasses the complexity of tests of the Gustafson model in the empirical literature.

Given linearity, high autocorrelation implies high persistence, in the sense that price next period is with high probability close to current price, whether or not the latter is far from the sample mean. A number of papers have confirmed the failure of storage arbitrage in a model with i.i.d. disturbances to match the high autocorrelations and/or coefficients of variation of observed prices. Others reach different conclusions on this issue. All of these papers, including those of which some among us are co-authors, accept the high reported price autocorrelations at face value, and acknowledge the need for their models to match these autocorrelations.

Figure 1, however, embodies a paradox: the fact that price is equally persistent at all price levels, high and low, does not imply that forward returns are independent of current price. The latter inference can be an illusion generated by the very trends that seem so negligible.

In Section I, after briefly describing our illustrative cotton and maize price data, we present nonparametric analyses of these price series. Consistent with Figure 1, they strongly imply a linear autoregressive process with high autocorrelation. The illustrated correspondences seem to support the case that investment is equally risky at high and low prices.

However when modest trends in these samples are appropriately recognized, an otherwise nonparametric analysis reveals correspondences displaying nonlinear stochastic behavior consistent with intertemporal arbitrage of stocks, subject to a non-negativity constraint.

Accordingly we present in Section II a nonstationary model of a market with trending stochastic production and intertemporal arbitrage of stocks, subject to a non-negativity constraint. In this model the trend does not affect the expected return on any stocks held, making neglect of the trend, at least by those who focus on near-term price spreads, more understandable.

However, whether expressed in terms of trending or detrended prices, our model must account for the effect of the trend on arbitrage activity. Studies that estimate parameters of models involving storage arbitrage (including Deaton and Laroque 1992, 1995, 1996, Miranda and Rui 1999, Osborne 2004, Roberts and Schlenker 2013, Cafiero et al. 2011, 2015, and Gouel and Legrand 2016), if they recognize the existence of trends, do not address their effects on storage arbitrage, nor do they prove consistency of their estimators.

---

1 Subsequent contributions include Samuelson (1971), Gardner (1979), Scheinkman and Schechtman (1983), Wright and Williams (1984), Deaton and Laroque (1992), and Bobenrieth, Bobenrieth and Wright (2002).


4 Bobenrieth, Wright, and Zeng (2013) includes the effects of the trend in the consistent estimation procedure underlying the published results but does not present the methodological details. Notes on the latter were prepared by Bobenrieth and Bobenrieth (2010).
The implications of this model for the correspondences relating prices and returns are complex. Simulations of a heuristic example of our model, in Section III, sharpen inferences from the nonparametric analysis in Section I, with respect to the state dependence and skewness of returns. They support the inference that the two price samples show evidence of storage arbitrage.

We present a method of estimation of key parameters of this model in Section IV, using a very general demand specification. Our proof of consistency of the estimators is relevant for estimation of nonlinear empirical models in which the long run trend is not known to be zero, and decisions obey Euler equations with occasionally binding constraints.

Section V presents estimates of three key parameters, the trend, the interest rate and the threshold price at which stocks go to zero, using our samples of cotton and maize prices. Concluding remarks follow.

I. A naive empirical exploration of two commodity price series

A. Data

Our two illustrative price series are samples of annual cotton and maize prices. We choose cotton because it is one of the industrial commodities included in Cashin and McDermott (2002), and the commodity referenced by Deaton and Laroque (1996 pp. 913 and 914) as exemplifying the high persistence of price at all price levels in the sample interval 1900-1987. We add maize because it is a major traded agricultural commodity with clearly trending price and yield, and plausibly little interaction with the global cotton market. Cotton and maize yields show average annual increases that do not appear to change with the level of yield over the period 1960-2007.

We chose the sample interval (1960-2007) to avoid demand shifts associated with large wars which (as Kindleberger recognized) can induce persistent price behavior atypical of peacetime markets, and to minimize complications associated with multiple changes in trend that might affect many commodity price series in the longer interval, 1900-1987, used in Cafiero et al. (2011) and earlier papers. Perhaps because we avoid these complications, our sample furnishes a more convincing case for the persistence of price at all price levels, and indeed the linearity of the autoregression correspondence, than does the earlier and longer price series for cotton. The sample interval excludes the major effects of the persistent and unprecedented boost in maize demand related to United States mandates and subsidies for biofuels, with implementation starting around 2007, and of recently fast-rising Chinese imports of maize and soybeans.

The nominal prices are from Pfaffenzeller, Newbold, and Rayner (2007) for 1960-2003, extended to 2004-2007 using Pfaffenzeller (2013). We take their annual calendar

\footnote{Kilian (2009) has noted the potential of unobserved shifts in expectations of wars to generate non-linearities in the behavior of the petroleum market.}
averages, and then normalize them by the 1977-79 average, following the description given in Pfaffenzeller, Newbold, and Rayner (2007). The deflator for the series is, except where otherwise indicated, the Manufactures Unit Value (MUV) Index used in many papers in the Prebisch-Singer tradition.\(^6\)

**B. Forward price differences: a nonparametric exploration**

Figures 2 and 3 present mappings of forward log price differences to current log prices, for cotton and maize respectively.

The top panels map the one-year forward differences in log prices to observed log prices. These correspondences, like Figure 1, suggest that the relative range of variation after any given price is not sensitive to the level of the latter.

Due to the dispersion of production, the one-year differences might well be contaminated by averaging of information related to different “crop years” in different parts of the world, smoothing spikes and, as first noted by Working (1960), inducing spurious autocorrelation. The bottom panels show two-year differences, less contaminated by this problem. Both panels support the inference drawn from Figure 1 of price persistence and price variation independent of current price. Indeed Figures 1-3 offer little reason to question the inference that the price autoregression function does not flatten out at high observed price levels, and that the distribution of forward returns is independent of current price.

**C. Implications of small trends for nonparametric inferences**

We now extend the above exploration, making minimal use of an estimated time trend. Given the size of our samples we do not pretend that we can present a definitive analysis of the price dynamics of either commodity. We consider a simple loglinear price trend. If there is any such trend in cotton and maize prices, it will be a constant displacement of the mean log difference, which one would expect to be manifest as a shift in the sample mean in Figures 2-3. These figures suggest that, if such a shift exists, its annual value must be negligible relative to the typical range of year-to-year price variation, as implied by the quote from Cashin and McDermott above, and as often assumed in recent empirical studies for agricultural producers and businesses such as Zulauf, Rettig, and Roberts (2014). Indeed, the trends in price estimated from regression of log prices on a linear time trend using ordinary least squares are −2.9% and −2.2%, whereas the standard deviations of log price changes are 16.7% and 16.2% for cotton and maize, respectively.

Figures 4 and 5 show the same log price differences on their vertical axes as those shown in Figures 2 and 3. But now these log price changes are mapped to base prices detrended by the trends reported in Table 1. This reordering of these observed log price differences reveals correspondences that appear to be declining in the detrended base price when the latter is high.

---

\(^6\)The samples from 1900-1987 referenced above are deflated by the United States CPI; the MUV does not cover the early part of that interval.
Figure 2: Cotton: log price mapped to one-period and two-period changes in log price
Figure 3: Maize: log price mapped to one-period and two-period changes in log price
Figure 4: Cotton: log detrended price, mapped to one-period and two-period changes in log price
Figure 5: Maize: log detrended price, mapped to one-period and two-period changes in log price
Table I: Trend estimates and their implications for sample moments

<table>
<thead>
<tr>
<th>Trend estimates</th>
<th>Prices</th>
<th>AC1</th>
<th>AC2</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cotton</td>
<td>-2.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observed Prices</td>
<td>0.93</td>
<td>0.86</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Detrended Prices</td>
<td>0.54</td>
<td>0.14</td>
<td>0.17</td>
</tr>
<tr>
<td>Maize</td>
<td>-2.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observed Prices</td>
<td>0.90</td>
<td>0.78</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Detrended Prices</td>
<td>0.63</td>
<td>0.21</td>
<td>0.21</td>
</tr>
</tbody>
</table>

For example, for cotton consider the ten highest detrended prices on the horizontal axes of Figure 4. For each of these, the top panel shows that the price falls the following year. The bottom panel shows that for each of these ten highest prices, the price fall after two years, less affected by problems of seasonal overlap, is more than 14% for each of these ten highest detrended prices.

These patterns contrast with the strong persistence of trending prices at all observed price levels implied by Figures 1, 2 and 3. When the small trends are taken into account, the occurrence of locally high prices as occasional large spikes interspersing longer intervals of less volatile prices, clearly evident in the graphs of the price data (see Figures 10 and 11 below), emerges in the autoregression correspondence and related representations of the dynamics of trending cotton and maize prices.\(^7\)

The intuition for the dramatic change induced by detrending of base prices in these correspondences is as follows: high detrended prices are not persistent, but the mapping from detrended prices to trending prices is not one to one. Low trending prices can, for example, correspond to a low detrended price at the beginning of the sample, or to a much higher detrended price at a later date in the sample. Thus the order induced by trending prices on the horizontal axes of Figures 2 and 3 mixes together observed prices of similar magnitude, mapped from very different detrended prices, with very different distributions of forward returns.

Table 1 shows that when prices are detrended using trend estimates in the table, first order autocorrelation drops from 0.93 to 0.54 for cotton, and from 0.90 to 0.63 for maize. Second order autocorrelation drops even more drastically, from around 0.8 to around 0.2 or less. Relative variability as indicated by the sample coefficient of variation, drops from 0.43 to 0.17 for cotton, and from 0.36 to 0.21 for maize.\(^8\)

\(^7\)Note that the apparent lack of price observations around the middle of the log price series in Figures 1, 2 and 3 is not replicated in Figures 4 and 5. As evident in Figures 10 and 11 below, there are several price observations close to the loglinear trend line.

\(^8\)How can a modest, unrecognized percentage trend induce such overestimation of autocorrelation in price time series? The sample autocorrelation is a sum of the products of the distances of successive sample observations from their respective means, normalized by the product of their standard deviations. Take any sample of price observations with an unrecognized negative percentage trend. Consider any two successive price observations in the sample. After the sample size is sufficiently extended, these two observations will both become located above the mean, as the mean drifts down and away from both, following the trend towards zero. Thus the estimated autocorrelation will tend to rise as the sample interval is extended forward. Estimated volatility also tends to increase, by a similar rationale. Indeed if the price mean approaches zero, any constant negative percentage trend implies that the sample
The dynamic price patterns observed in our samples, and the effects of their estimated trends, are independent of any assumption about the operation of storage arbitrage or the nature of the price behavior, apart from the assumed loglinear trend. Note, however, that the apparent tendency of the log observed price changes in Figures 4 and 5 to be increasingly negative as detrended prices rise is suggestive of the expected behavior of price changes from high prices in the Gustafson model.

On the other hand, inspection of Figures 4 and 5 might well raise the question of whether the correspondences might be linear, with expected returns declining at all log detrended prices, in contrast to the storage model, which implies that expected relative price changes are constant for prices below the trending price threshold. We defer exploration of this question to Section III, where we consider a similar figure (Figure 7) generated by a heuristic numerical model of competitive storage arbitrage for a commodity with a deterministic trend in price.

II. Trends and storage arbitrage

A. The model

For many important commodities, long run global price declines are commonly attributed to the effects of persistently increasing productivity, which we specify to be such that it implies the following deterministic trend in the price $P_t$:

$$P_t = \lambda t p_t,$$

where $\lambda > 0$ and $p_t$ is the detrended price. Prices $P_t$ and $p_t$ are given by $P_t = F(C_t)$ and $p_t = F(c_t)$, where $F$ is the inverse consumption demand, and $C_t$, $c_t$ denote consumption and detrended consumption, respectively.\(^9\)

Assume that $F$ is stationary, strictly decreasing, and satisfies:

$$\frac{F''(F)}{(F')^2} = \kappa,$$

where $\kappa$ is a constant.\(^{10}\) Integrating (2) we obtain $F(C) = (A + BC)^{1/\kappa}$ if $\kappa \neq 1$, and $F(C) = e^{A+BC}$ if $\kappa = 1$, where $A$ and $B$ are constants. Thus our specification includes linear, log-linear, and iso-elastic inverse consumption demand functions as particular cases, given appropriately chosen values for $A$ and $B$.

Assume that detrended production $h_t$ is i.i.d., and has compact support $[h, \overline{h}]$, where $-\infty < \underline{h} < \overline{h} < \infty$. Assume $EF(h) < \infty$, where $E$ denotes the expectation with respect to $h$.

\(^9\)For a model in which oil demand is a function of trending income, see Dviv and Rogoff (2009, 2014).

\(^{10}\)Thus, $F$ has the form of a derivative of a Hyperbolic Absolute Risk Aversion (HARA) function. For a detailed discussion of HARA functions, see the Appendix in Carroll and Kimball (1996).
Begin with the case in which storage is by assumption infeasible. In this case consumption $C_t$ equals production $H_t$, detrended consumption $c_t$ equals detrended production, $h_t$, and therefore detrended price is $p_t = F(h_t)$. Equation (1) then implies a deterministic trend in production $H_t$ implicitly defined by the following equation:

$$F(H_t) = \lambda^t F(h_t).$$

(3)

Equation (3) implies that $E_t P_{t+1} = E_t F(H_{t+1}) = \lambda^{t+1} EF(h_t)$, where $E_t$ denotes the expectation conditional on information at time $t$.

We now allow for non-negative stocks. We assume no irreducible “pipeline” or “working” stocks essential for operation of the market. We also assume there is no storage cost apart from a constant interest rate $r > 0$. Competitive expected profit maximization implies that, when stocks are positive,

$$E_t P_{t+1} = (1 + r) P_t,$$

thus the percentage spread between the expectation of price in the next period and current price is $r$, regardless of the production trend. However, the assumption of bounded $EF(h_t)$ implies the occurrence of a stockout in finite time.

Let $Z_t$ denote available supply at time $t$. Price obeys the following condition:

$$F(C_t) = \max \left[ F(Z_t), \frac{1}{1 + r} E_t F(C_{t+1}) \right], \text{ s.t.}$$

$$Z_{t+1} = Z_t - C_t + H_{t+1}, \forall t \in \mathbb{N}.$$ 

(4)

(5)

This model has a normalized representation, with a stationary rational expectations equilibrium. Equations (4) and (5) have the following normalized counterparts, with detrended available supply $z_t$ implicitly defined by $F(Z_t) = \lambda^t F(z_t)$:

$$F(c_t) = \max \left[ F(z_t), \frac{\lambda}{1 + r} E_t F(c_{t+1}) \right], \text{ s.t.}$$

$$z_{t+1} = \lambda^{\kappa-1} (z_t - c_t) + h_{t+1}.$$  

(6)

(7)

Assuming $\lambda < 1 + r$, standard arguments imply the existence of a stationary rational expectations equilibrium price function $p$ for the detrended price:

$$p_t = p(z_t) = \max \left[ F(z_t), \frac{\lambda}{1 + r} E_t p(z_{t+1}) \right].$$

(8)

The logic of the proof of Theorem 1 of Deaton and Laroque (1992) implies that $p$ is non-negative, continuous, strictly decreasing, and that the following complementary inequalities hold:

11Equation (7), is obtained from (5) using the following property of $F$: for any $K > 0$ and $y \in \mathbb{R}$, that allows $F$ to be well defined, $F^{-1} (KF(y)) = [K^{1-\kappa} (By + A) - A] / B$ if $\kappa \neq 1$, and $F^{-1} (KF(y)) = \ln(K) / B + y$, if $\kappa = 1$. 

13
\[ p(z) = F(z), \text{ for } z \leq F^{-1}(p^*), \]
\[ p(z) > F(z), \text{ for } z > F^{-1}(p^*), \]

where \( p^* \equiv \left( \frac{\lambda}{1 + r} \right) Ep(h) \in \mathbb{R}. \)

Equation (8) implies the following autoregression for detrended prices:

\[ E_t p(z_{t+1}) = \left( \frac{1 + r}{\lambda} \right) \min[p^*, p(z_t)]. \tag{9} \]

Multiplying (9) by \( \lambda^{t+1}, \) we obtain the autoregression expressed in terms of observable prices:

\[ E_t P_{t+1} = (1 + r) \min[\lambda^t p^*, P_t]. \tag{10} \]

B. Spreads on stocks held: detrending reveals the trends

The trend (represented by \( \lambda \)) appears in the expectation of the price (equation (10)) only in the expression for the current threshold price. If current price \( P_t \) exceeds the trending threshold \( \lambda^t p^* \), then the “spread” between the conditional expectation and the current price, as a percentage, \( \frac{E_t P_{t+1} - P_t}{P_t} \), reflects the current price, the current threshold and the interest rate. That is, when price is at a spike, the spread indicates an expected jump to \( (1+r)\lambda^t p^* \). By contrast, when stocks are positive, the spread is independent of the trend and equal to the interest rate, \( r \). Observation of such dynamic behavior might encourage an incorrect inference that the underlying trend is a fundamental jump process in price. Actually, conditional on current trending price and on the current threshold price, the trend never affects the spread, whether or not stocks are positive.

Ironically, the trend always affects the spread expressed in detrended prices, whether or not conditional on the current price and the current threshold. If the model has no trend (\( \lambda \equiv 1 \)), the threshold \( p^* \) is higher and stocks are larger than if \( \lambda < 1 \). Hence to replicate the storage activity in the trending model, the opportunity cost of storing in the Euler equation expressed in detrended prices must be increased, multiplying by \( 1/\lambda \) (see equation (9)).

III. A heuristic exploration: Are small trends negligible?

To explore the effects of a negative price trend, we conduct Monte Carlo experiments with a heuristic model with the specification introduced in Wright and Williams (1982), and re-used in Williams and Wright (1991, pp. 58-62), Deaton and Laroque (1992, p. 11) and Cafiero et al. (2011, pp. 45-46). Inverse consumption demand has the specification
$F(C) = 600 - 5C$, harvest has a normal distribution with mean 100 and standard deviation 10, and the only cost of storage is the interest rate, $r = 0.05$. Deaton and Laroque (1992) show that this example implies price autocorrelations too small to match those they observe for major commodities. It also produces a coefficient of variation too small to match the values they report for their commodity price data, given storage is allowed.

We alter this heuristic model only by specifying a price trend of minus two percent per year (of the order of magnitude of the trends we estimate for cotton and maize), and implement a numerical simulation of the model described in the previous section, taking account of the effect of the trend on arbitrage incentives. We truncate the normal distribution at plus and minus 5 standard deviations from the mean, as in Deaton and Laroque (1992) and Cafiero et al. (2011), numerically generate prices for 300,000 time periods, and take successive samples of size 48 (the same sample size we use in our empirical estimations) from the simulated series, the first starting from period $t = 1$, the second from period $t = 2$, and so on. For samples of this size, Dickey Fuller tests at the 5% rejection level fail to reject the unit root null in 81% of log price autoregressions. If time is included in the autoregression the failure rate is 21%.

Figure 6, based on samples of 48 prices starting at $t = 1$, $t = 48$, etc. (with each initial value normalized to its detrended level),\textsuperscript{12} shows the correspondence between the log of current trending price and its first difference in these samples. The apparent linearity of the correspondence in Figure 6, and the apparent lack of dependence on current price, are strongly reminiscent of the nonparametric correspondences in Figures 2 and 3. They imply a linear upsloping autoregression correspondence like Figure 1, and are consistent with empirical claims that commodity prices are highly persistent even during booms.

Figure 7, generated from 299,999 couples of current price and subsequent price in the simulated string of 300,000 prices, shows the correspondence between the log of detrended current price and the same first differences of the logs of trending prices shown in the previous figure. The dark patch of observations indicate that most of the mass in the invariant distribution is concentrated to the left of the threshold base detrended price. We can use Figure 7, and what we know about the underlying dynamic stochastic model, to refine inferences from nonparametric analyses of observed prices and returns, as presented for example in Figures 4 and 5 above. The downslope of the dark patch might suggest that mean conditional returns are decreasing and linear in log detrended base price to the left of the threshold, and indeed for all log detrended prices, an inference that might well be reinforced by inspection of Figures 4 and 5.

However this inference is incorrect. In Figure 7, for every log detrended price to the left of the log of the threshold price $p^*$, stocks are positive, and the expected return to storage with trending prices is constant at the interest rate. At the lowest detrended price, price changes in this heuristic model are positive and roughly symmetric around the expected change, so the changes in log prices are approximately symmetric too. At higher log detrended prices, the distribution of these changes in log prices becomes increasingly skewed upward as detrended price rises toward the threshold and the stocks available to

\textsuperscript{12}This sampling procedure ensures that each of the first 299,999 prices is considered only once in the horizontal axis.
Figure 6: Heuristic model with trend: correspondence between the log price $\ln P_t$ and first difference of log price $\ln P_{t+1} - \ln P_t$.

Figure 7: Heuristic model: correspondence between the log detrended price $\ln p_t$ and first difference of log price $\ln P_{t+1} - \ln P_t$. 
cushion the effects of a bad harvest decline; the concentrated mass in Figure 7 must trend down to maintain the constant expected price change.

The light vertical flare above the horizontal line and to the left of the vertical dashed line indicates the prevalence of price jumps from prices close to, but below the stockout threshold. Given \( t \), for price realizations higher than the stockout price (to the right of the vertical dashed line), the expectation of the subsequent price is constant at \( (1 + r)\lambda p^t \). Conditional on this expectation, the expected drop increases one-for-one with current price. That is why the correspondence appears to have a downslope of about 45 degrees to the right of the threshold in Figure 7.

Together, Figures 6 and 7 comprise a dramatic illustration of the empirical mischief that can be made by a neglected modest trend in price, even in a relatively small sample. The trend transforms the complex nonlinear relation between detrended price and the distribution of immediately subsequent return realizations, illustrated by the correspondences in Figure 7, into the picture of a constant distribution of return realizations, independent of base trending prices, shown in Figure 6. In this heuristic model, high prices are, as Kindleberger warned, likely to be fleeting. Repeated high prices (in the top right hand quadrant of Figure 7) are sparse.

The effects on estimated moments of removing the trend, presented in Table 2 and Figure 8, are also striking. The median of the sample first order autocorrelation coefficients falls from 0.71 to 0.41, the median second order autocorrelation falls from 0.60 to 0.21 and the median of the sample coefficient of variation falls from 0.38 to 0.26. These effects on autocorrelation, noted in Gouel and Legrand (2016), are not small sample phenomena, indeed they increase with the sample size. At a sample size of 3,000, the medians of the sample first and second-order autocorrelation coefficients of trending prices are close to 1, and the median of the sample coefficient of variation for trending prices is about 20 times that of the detrended prices. At this sample size of 3,000, even a trend of \(-0.02\) percentage points per year dramatically increases estimated autocorrelation and coefficient of variation, and induces linearity of spreads in observed price.\(^\text{13}\)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Prices</th>
<th>AC1</th>
<th>AC2</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>Trending</td>
<td>0.71</td>
<td>0.60</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Detrended</td>
<td>0.41</td>
<td>0.21</td>
<td>0.26</td>
</tr>
<tr>
<td>3,000</td>
<td>Trending</td>
<td>0.96</td>
<td>0.95</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td>Detrended</td>
<td>0.44</td>
<td>0.27</td>
<td>0.28</td>
</tr>
</tbody>
</table>

In the context of intertemporal accumulation detrending reorders sample prices ac-

\(^{13}\)If there is actually no price trend (i.e. \( \lambda \equiv 1 \)), and we remove an estimated trend before calculating the sample statistics, this induces much smaller biases in the first and second-order sample autocorrelation and sample coefficient of variation, given the sample interval of 48, than if we assume there is no trend when there is in fact a \(-2\%\) trend. These biases induced by inappropriate detrending decrease with sample size.
Figure 8: Empirical distributions from samples of size 48: AC1, AC2, CV. (Smoothing is based on a normal kernel.) The vertical lines indicate the corresponding medians according to their distance from the constant detrended threshold $p^*$, so that the correspondence between forward returns and detrended prices reveals the nonlinear relation implied by constrained storage arbitrage. Neglect of a small trend or similarly persistent influence on price can hide the fact that when the price is above the trending threshold, the spread is a jump equal to the difference between current price and $(1 + r)$ times that threshold. Given that trending price is a strictly increasing function of detrended price, and a strictly decreasing function of time, implies that different values for trending price can generate identical returns. To take a simple example, consider a sequence of detrended prices, all equal to $p$. In this case, the corresponding sequence of trending prices is $\{\lambda t p\}_{t \in \mathbb{N}}$, strictly decreasing with time, and the corresponding realized returns are $\frac{P_{t+1} - P_t}{P_t} = \lambda - 1$, identical for all $t$.

Our numerical results make a case for developing an econometric procedure that does not impose the assumption that the trend is exactly zero. However, without this assumption the question of consistency of the estimators naturally arises. In the next section we present our proofs of superconsistency in probability for the least squares estimator of the trend in a first step regression, and consistency in probability for (second step) nonlinear least squares estimators of $r$ and $p^*$.

IV. Consistency of least squares estimators for $\lambda$, $r$, and $p^*$

We follow the empirical literatures in the traditions of Prebish-Singer or Gustafson
in focusing only on price data, which are widely considered to be more accurate than quantity data in global commodity markets. Our focus is on estimation of the trend parameter, $\lambda$, the interest rate, $r$, and the detrended threshold price, $p^*$, three key parameters of the model. We consider the following two step procedure. In the first step estimate the trend using least squares on a regression of log prices on time and a constant. In the second step, given the trend estimated in the first step, estimate $r$ and $p^*$ using nonlinear least squares on a normalized representation of (10).

Let $\gamma \equiv 1 + r$. For clarity of exposition, in the remainder of this paper a subscript 0 on a parameter denotes its true value.

Equation (1) can be re-written in natural logarithms as:  
\[
\ln P_t = t \ln \lambda_0 + \ln p_t. 
\]  
(11)

Under conditions that guarantee the existence of a unique invariant distribution for detrended prices, define $\alpha_0 \equiv E_\infty [\ln p]$, where $E_\infty$ denotes the expectation with respect to the invariant distribution of detrended prices. Assume that the minimum detrended price in the ergodic set is strictly positive. Hence $\alpha_0$ is a finite number. Then we can rewrite (11) as:  
\[
\ln P_t = \alpha_0 + \beta_0 t + \xi_t, 
\]  
(12)

where $\beta_0 \equiv \ln \lambda_0$ and $\xi_t \equiv \ln p_t - \alpha_0$.

The least squares estimator $\hat{\beta}_T$ of $\beta_0$ satisfies:  
\[
T \left( \hat{\beta}_T - \beta_0 \right) = \frac{-6T}{T - 1} \left[ \frac{1}{T} \sum_{t=1}^{T} \xi_t \right] + \left[ \frac{12T^2}{T^2 - 1} \right] \left[ \frac{1}{T^2} \sum_{t=1}^{T} t \xi_t \right]. 
\]  
(13)

In classical linear trend models with uncorrelated and bounded disturbances it is straightforward to prove that the least squares estimator of the time trend coefficient is superconsistent. Note that the disturbances $\xi_t$ in equation (12), which are deviations of detrended log price from its long run mean are not uncorrelated and can be unbounded.

We now prove superconsistency in probability of $\hat{\beta}_T$. For clarity of exposition we list our assumptions:

Assumption 1: $\{h_t\}_{t\in\mathbb{N}}$ is i.i.d., with bounded and compact support $[\underline{h}, \overline{h}]$, and a continuous distribution with strictly positive derivative in $(\underline{h}, \overline{h})$.

Assumption 2: $EF(h) < \infty$.

Assumption 3: $p_t \geq \underline{p} > 0$, for all $t \in \mathbb{N}$.\footnote{For linear consumption demand, Bobenrieth and Bobenrieth (2010) provide sufficient conditions for the existence of $\underline{p} > 0$ with $p_t \geq \underline{p}$, for all $t$.}

19
Theorem 1: Let Assumptions 1, 2, and 3 hold. Then, \( T \left( \hat{\beta}_T - \beta_0 \right) \to 0 \), in probability. Therefore \( \hat{\lambda}_T \equiv e^{\hat{\beta}_T} \) satisfies:

\[
\left[ \hat{\lambda}_T / \lambda_0 \right]^T \to 1, \text{ in probability.}
\]

The proof of Theorem 1 is presented in Appendix A. We prove that the two terms on the RHS of (13) go to zero. We first note that by the strong law of large numbers of Breiman (1960), the existence of a unique invariant distribution for detrended prices implies that, almost surely, \( \frac{1}{T} \sum_{t=1}^{T} \xi_t \to 0 \). To show that \( \frac{1}{T^2} \sum_{t=1}^{T} t \xi_t \) goes to zero in probability, we use the facts that the detrended price process is irreducible and its corresponding Markov operator is stable and quasicompact, concluding that the sequence of log detrended prices is uniformly integrable, uniformly in initial detrended price. We then apply a weak law of large numbers for triangular arrays due to Andrews (1988, Theorem 2(a)).

Autoregression (10) can be written as:

\[
P_{t+1} = \gamma_0 \min \{ \lambda_t^0 p_t^*, P_t \} + \zeta_{t+1}, \quad \text{where} \quad E_t(\zeta_{t+1}) = 0.
\] (14)

Given \( \hat{\lambda}_T \), estimated in the first step, we estimate \( p_t^* \) and \( \gamma_0 \) using nonlinear least squares.

Note that if \( \lambda_0 < 1 \) we cannot identify \( (\gamma_0, p_0^*) \) in (14). Indeed, for \( (\gamma_0, p_0^*) \neq (\gamma_0, p_0^*) \), and \( f_t((\gamma, p^*), \lambda) \equiv \gamma \min \{ \lambda^t p^*, P_t \} \), there exists a ball \( B(\gamma_0, p_0^*) \) centered at \( (\gamma_0, p_0^*) \) such that the accumulated one-period ahead squared predictive bias, \( \sum_{t=1}^{T} \left\{ f_t((\gamma, p^*), \hat{\lambda}_T) - f_t((\gamma_0, p_0^*), \lambda_0) \right\}^2 \), satisfies:

\[
\inf_{(\gamma, p^*) \in B(\gamma_0, p_0^*)} \sum_{t=1}^{T} \left\{ f_t((\gamma, p^*), \hat{\lambda}_T) - f_t((\gamma_0, p_0^*), \lambda_0) \right\}^2 \leq \sum_{t=1}^{+\infty} \lambda_t^0 \varrho < \infty,
\]

where \( \varrho < \infty \).

To avoid this problem without assuming \( \lambda_0 \geq 1 \), we divide the regression model (14) by \( P_t \), to obtain the following:

\[
\frac{P_{t+1}}{P_t} = \gamma_0 \min \left\{ \lambda_t^0 p_t^*, 1 \right\} + \varepsilon_{t+1},
\] (15)

\( ^{15} \)For discussions of identification in models characterized by a martingale difference sequence, see for example Wu (1981) and Skouras (2000, Remark 1, p. 873).
where $\varepsilon_{t+1} \equiv \frac{\zeta_{t+1}}{P_t}$ satisfies $E_t(\varepsilon_{t+1}) = 0$.

Let $g_t$ be the corresponding predictor, that is:

$$g_t((\gamma, p^*), \lambda) \equiv \gamma \min \left\{ \lambda^t \frac{p^*}{P_t}, 1 \right\} = \gamma \min \left\{ \left( \frac{\lambda}{\lambda_0} \right)^t \frac{p^*}{p_t}, 1 \right\}, \text{ for } t \in \mathbb{N}.$$  

For any given sample size $T \in \mathbb{N}$, define $(\hat{\gamma}_T, \hat{p}^*_T)$ to be the least squares estimators for $(\gamma_0, p^*_0)$, that is:

$$(\hat{\gamma}_T, \hat{p}^*_T) \equiv \arg\min_{(\gamma, p^*)} \frac{1}{T} \sum_{t=1}^{T} u^2_t((\gamma, p^*), \hat{\lambda}_T),$$

where

$$u_t((\gamma, p^*), \hat{\lambda}_T) \equiv \frac{P_{t+1}}{P_t} - g_t((\gamma, p^*), \hat{\lambda}_T).$$

To prove consistency (in probability) of $(\hat{\gamma}_T, \hat{p}^*_T)$, we now add the following conditions:

Assumption 4: $EF^2(h) < \infty$.

Assumption 5: $p^*_0 > p$.

Assumption 6: The parameter space is compact.

**Theorem 2:** Let Assumptions 1, 2, 3, 4, 5, and 6 hold. Then, $(\hat{\gamma}_T, \hat{p}^*_T) \rightarrow (\gamma_0, p^*_0)$, in probability.

**Proof of Theorem 2:** Appendix B.

**Remark:** The residual of the first step regression, which is the deviation of log detrended price from its invariant mean, is ergodic, and is a nonlinear function of $r$ and $p^*$, which are parameters estimated in the second step regression.

### V. Empirical results

Before discussing the empirical implementation with price data, we address the small sample performance of our estimator. We use the heuristic model in section III, and include all possible samples of 48 consecutive prices drawn from a numerically generated series of 300,000 prices. Consider first the case where we specify that there is no trend in the model, i.e. $\lambda_0 \equiv 1$. In this setting our econometric procedure consists only of the second-step estimators, for $r_0$ and $p^*_0$. For this case, Table 3 shows that the performance of our least squares estimator is similar to that of the Generalized Method of Moments (GMM) of Deaton and Laroque (1992). For the case with $\lambda_0$ set at 0.98, using our least squares procedure, the mean and the median of the estimated $\lambda$ are both 0.98. The mean
Table III: Numerical Performance of Estimators\textsuperscript{†}

<table>
<thead>
<tr>
<th></th>
<th>Least Squares</th>
<th>GMM\textsuperscript{‡}</th>
<th>Two-Step Least Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0=1$, $r_0=0.05$, $p^*_0=113.91$</td>
<td>$\lambda_0=1$, $r_0=0.05$, $p^*_0=113.91$</td>
<td>$\lambda_0=0.98$, $r_0=0.05$, $p^*_0=109.46$</td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>$p^*$</td>
<td>$r$</td>
<td>$p^*$</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0655</td>
<td>109.56</td>
<td>0.0654</td>
</tr>
<tr>
<td>Median</td>
<td>0.0587</td>
<td>109.15</td>
<td>0.0577</td>
</tr>
<tr>
<td>St. dev.</td>
<td>0.0393</td>
<td>16.16</td>
<td>0.0412</td>
</tr>
<tr>
<td>Root Mean</td>
<td>0.0423</td>
<td>16.74</td>
<td>0.0440</td>
</tr>
<tr>
<td>Square Error</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{†}We drop from the Monte Carlo the estimates of $p^*$ larger than or equal to the maximum price in the sample. In those cases (1.27%, 0.45%, and 0.73%, of the total number of samples considered, for least squares with $\lambda = 1$ and $\lambda = 0.98$, and for GMM, respectively) the minimization of the objective function cannot identify $p^*$.

\textsuperscript{‡}Generalized Method of Moments estimator for a stationary storage model (Deaton and Laroque 1992).

and median of the estimated $r$ are similar to those obtained if there is no trend in the model (i.e. if $\lambda_0 \equiv 1$).

We implement our two step empirical procedure using the annual price data for cotton and maize discussed in Section I above. In the first step of our estimation procedure, we estimate the percentage trend in each annual price series, with results presented in Table 1.

In the second step, using nonlinear least squares, we estimate the interest rate $r_0$ and the threshold price $p^*_0$, given the point estimates of the trends, $-2.9$ and $-2.2$ percent for cotton and maize, respectively. The results are presented in Table 4. Ignoring the negative trend in the Euler equation induces a negative bias in the estimate of the interest rate, so large in the case of cotton that it becomes negative. When the trend is recognized, the interest rates are remarkably similar at 2.92 percent and 2.74 percent for cotton and maize, respectively.

Table IV: Estimates of $r_0$ and $p^*_0$

<table>
<thead>
<tr>
<th></th>
<th>$r$</th>
<th>$p^*$</th>
<th>% of stockouts in the sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cotton</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ignoring Trend</td>
<td>-1.08%</td>
<td>0.61</td>
<td>4.17%</td>
</tr>
<tr>
<td>Recognizing Trend</td>
<td>2.92%</td>
<td>0.64</td>
<td>43.75%</td>
</tr>
<tr>
<td>Maize</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ignoring Trend</td>
<td>0.90%</td>
<td>0.64</td>
<td>4.17%</td>
</tr>
<tr>
<td>Recognizing Trend</td>
<td>2.74%</td>
<td>0.73</td>
<td>16.67%</td>
</tr>
</tbody>
</table>

In Figure 9 we plot, as in the first panels of Figures 4 and 5, the logarithm of each detrended price in the sample (horizontal axis), and the first difference of the logs of prices (vertical axis), for cotton and maize, but we superimpose on the correspondence the function implied by the estimates of $\lambda_0$, $r_0$, and $p^*_0$.  

22
Figure 9: Cotton and Maize: estimated functions relating log detrended price to one-period changes in log price
Figure 10: Real cotton prices, real price threshold, and the price predicted by the model

Figure 11: Real maize prices, real price threshold, and the price predicted by the model
The estimated functions are constant to the left of the threshold price, but then slope down at 45 degrees, as suggested in the correspondence generated by the heuristic model, illustrated in Figure 7 above.

When the trend is recognized, for our samples from 1960 to 2007. Figure 10 shows twenty-one stockouts for cotton, but only a handful of prominent price spikes. Figure 11 indicates that there are eight stockouts for maize.\footnote{Cafiero et al. (2011), assuming no trend, found no stockouts in a sample of maize prices from 1900-1987, a result that implies very high accumulation of stocks in the last years of their sample. In contrast, using our two-step procedure we find thirteen stockouts in that sample, assuming one constant secular trend over the entire eighty-eight year period. (Results are available from the authors.)}

Local peak prices appear to be trending down over time. Compared to any price below the threshold, a stockout price is more likely to occur next period if the current price is a stockout price. The shock that produces a locally first stockout in period $t$ is damped by carried-in stocks. Subsequent shocks are directly realized one-for-one as cuts in consumption, undamped by carried-in stocks. This is one reason why the highest peaks above trend tend to occur after a prior stockout price, even if harvests are independent.\footnote{Another reason for one high price to follow another in these samples is the averaging of the price effects of one marketing year that overlaps two calendar years. (See Guerra et al. 2015.)}

Figures 10 and 11 show why estimates of price autocorrelation are not very informative regarding price persistence for storable commodities. Price persistence is very different, depending on whether or not stocks are positive. For current price above the threshold price $\lambda_t p^*$, the conditional expectation of price is, as noted above, independent of current price. The spread between expected price for next period and the current price decreases as the current price increases, dollar for dollar. Given any current price below the threshold price $\lambda_t p^*$, storage arbitrage is active and the percentage spread between the conditional expectation of price next period and current price is constant and equal to $r$, independent of the trend.

**VI. Concluding remarks**

Price trends that are small relative to annual price variation can transform inferences about behavior of commodity prices, linearizing highly nonlinear price autoregressions and generating sample returns that appear to be independent of current price. How can small trends have such dramatic effects? In our storage model there is a time-dependent price threshold $P^*_t$ above which stocks are zero. This threshold follows the trend, which we assume here to be negative. Consider a price realization $P$ early in a given sample, and assume it is below the trending threshold $P^*_t$. The distribution of the immediately subsequent price change has a mean of $rP$, consistent with active storage arbitrage. Consider, alternately, the case in which the same price realization $P$ occurs at time $t'$ later in the sample, when it is above the downtrending threshold, $P^*_{t'}$. In this case, with no stocks, the distribution of the subsequent price change has a mean of $P^*_{t'} - P$.

Hence, price changes from a given price $P$ observed sufficiently early and sufficiently late in the sample can have two distinct distributions. If the sample size is large enough,
trend-induced shuffling of observed prices with very different distributions of forward realizations will make return realizations appear to be approximately independent of current price, even though each price realization is a function of the current price, the harvest realization, and time.

Detrending reorders sample prices according to their distance from the constant detrended threshold \( p^* \), so that the correspondence between forward returns and detrended prices reveals the nonlinear relation, with a kink at \( p^* \), implied by constrained storage arbitrage.

In studies that recognize trends it is common practice to detrend the sample observations in a preliminary step, and then estimate the model relying on standard consistency proofs for models without trends as if they applied to the detrended model. We prove consistency in an empirical model that explicitly recognizes the influence of the trend on arbitrage, even after detrending, and illustrate its application in estimation of commodity market behavior.

Note that we do not wish to claim that our empirical estimates capture all the features of our illustrative samples. For example, one might well argue that cotton prices in Figure 10 seem to follow several different trends, steeply declining before 1970, then weakly declining until 1984, then almost flat in later years. A more accurate identification of secular trends, breaks in trends, stochastic trends, or persistent ergodic shifters, should only increase the force of our findings regarding the effects of ignoring such phenomena, in empirical analysis.

A similar approach might be useful in other applications, including estimation of models of precautionary saving with liquidity constraints and persistent variation in productivity or consumption demand.
References


Ng, S., and F.J. Ruge-Murcia. 2000. “Explaining the Persistence of Commodity
Prices.” *Computational Economics*, 16, 149–171.


Zulauf, C., N. Rettig, and M. Roberts. 2014. “Do Futures Forecast the Future?” Farmdoc Daily 4 (147), Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.
Mathematical Appendix

Appendix A. Proof of Theorem 1

We prove three preliminary results.

**Lemma A.1:** There exists \( t_1 \in \mathbb{N} \), \( t_1 \) independent of initial detrended price such that the sequence \( \{\ln p_t\}_{t \geq t_1} \) is uniformly integrable, uniformly in the initial detrended price. More precisely, given \( \varepsilon > 0 \), there exists \( M_\varepsilon > 0 \), \( M_\varepsilon \) independent of initial detrended price, with:

\[
M_\geq M_\varepsilon \Rightarrow \sup_{t \geq t_1} \int_{|\ln p| \geq M} |\ln p| \mu_t(dp) < \varepsilon,
\]
where \( \mu_t \) denotes the distribution of \( p_t \), conditional on initial detrended price.

**Proof of Lemma A.1:** Consider the non-trivial case where detrended prices are unbounded. It suffices to show that there exists \( t_1 \in \mathbb{N} \), and a constant \( K_1 \in \mathbb{R} \), \( t_1 \) and \( K_1 \) independent of initial detrended price, such that \( \sup_{t \geq t_1} E[(\ln p_t)^2] \leq K_1 \). Let \( \rho \equiv \max\{p_0^*, e\} \). Then, \( E[(\ln p_t)^2] = \)

\[
\int_{|\ln p| \geq \ln \rho} (\ln p)^2 \mu_t(dp) + \int_{|\ln p| < \ln \rho} (\ln p)^2 \mu_t(dp)
\]

\[
= \int_{\ln p \geq \ln \rho} (\ln p)^2 \mu_t(dp) + \int_{\ln p \leq -\ln \rho} (\ln p)^2 \mu_t(dp) + \int_{|\ln p| < \ln \rho} (\ln p)^2 \mu_t(dp)
\]

\[
\leq \int_{\ln p \geq \ln \rho} (\ln p)^2 \mu_t(dp) + (\ln \rho)^2 + (\ln \rho)^2.
\]

We now bound the first term in the RHS of the last inequality. Considering that the function \( p \mapsto (\ln p)^2 \) is concave in \([e, \infty[\), from Jensen’s inequality we have:

\[
\int_{p \geq \rho} (\ln p)^2 \mu_t(dp) \leq \left\{ \ln \left[ \frac{1}{\mu_t([\rho, \infty])} \int_{p \geq \rho} p \mu_t(dp) \right] \right\}^2.
\]

The measure \( \mu_t([\rho, \infty]) \) depends on the initial detrended price, and is non-decreasing in the initial detrended price. From the monotonicity of the storage function, we have that \( \mu_t([\rho, \infty]) \geq \mu_t([\rho, \infty]), \) where \( \mu_t \) denotes the conditional probability measure of detrended prices at time \( t \), conditional on initial detrended price equal \( p \).

Considering that \( p \mapsto (\ln p)^2 \) is increasing in \([e, \infty[\), and that:

\[
\int_{p \geq \rho} p \mu_t(dp) \leq (1 + r)p_0^*,
\]

31
we obtain the following bound, which does not depend on the initial detrended price:

\[ \int_{p \geq \rho} (\ln p)^2 \mu_t(dp) \leq \left\{ \ln \left[ \frac{1}{\mu_t([\rho, \infty])} (1 + r)p_0^* \right] \right\}^2. \]

Given that the distribution of the shocks \( \{h_t\}_{t \in \mathbb{N}} \) is continuous on a compact support, the detrended price process is irreducible and its corresponding Markov operator is stable and quasicompact. Using Theorem 3.6 in Futia (1982, p. 390), and Theorem 4 in Yosida and Kakutani (1941, p. 200), we conclude that the detrended price process has a unique invariant distribution, \( \mu_\infty \), which is a global attractor in the total variation norm. More precisely, there exist strictly positive constants \( L, \eta \) (which are independent of the initial detrended price), with:

\[ ||\mu_t - \mu_\infty|| \leq \frac{L}{(1 + \eta)^t}, \quad \forall \ t \in \mathbb{N}, \]

where \( || \cdot || \) denotes the total variation norm. From (16) we conclude that there exists \( t_1 \in \mathbb{N}, \ t_1 \) independent of initial detrended price, such that:

\( t \geq t_1 \Rightarrow \mu_t([\rho, \infty]) \geq \mu_\infty([\rho, \infty])/2 \). Therefore:

\[ \int_{p \geq \rho} (\ln p)^2 \mu_t(dp) \leq \left\{ \ln \left[ \frac{2}{\mu_\infty([\rho, \infty])} (1 + r)p_0^* \right] \right\}^2. \quad \text{Q.E.D.} \]

We now prove that the conditional expectation of log detrended prices converges uniformly on initial detrended price, to the expectation of log detrended prices with respect to its limiting distribution:

**Lemma A.2:** Given \( \varepsilon > 0 \), there exists \( t_2 \in \mathbb{N} \), such that:

\[ \left| E_0[\ln p_t] - E_\infty[\ln p] \right| < \varepsilon, \quad \forall \ t \geq t_2, \text{ for any initial detrended price } \]

\( \geq \rho \), \text{ for any initial detrended price } \geq \rho \).

**Proof of Lemma A.2:** For any given \( M > 0 \), we have:

\[ \left| E_0[\ln p_t] - E_\infty[\ln p] \right| = \left| \int_{|\ln p| \geq M} \ln p \ \mu_t(dp) + \int_{|\ln p| < M} \ln p \ \mu_t(dp) - \int_{|\ln p| \geq M} \ln p \ \mu_\infty(dp) - \int_{|\ln p| < M} \ln p \ \mu_\infty(dp) \right| \]

\[ \leq \left| \int_{|\ln p| \geq M} \ln p \ \mu_t(dp) - \int_{|\ln p| \geq M} \ln p \ \mu_\infty(dp) \right| \]
implies that, given array $\epsilon$, that fact that $\leq$, we conclude the proof. Q.E.D.

Lemma A.2, together with the homogeneity of the detrended price Markov process implies that, given $\epsilon > 0$, there exists $t_2 \in \mathbb{N}$ such that:

$$
\forall m \geq t_2, \; \forall t \geq m, \; \forall p_{t-m} \geq \underline{p} \quad \text{implying that:}
\begin{align*}
&\sup_{t \geq m} E \left\{ E \left[ \ln p_t \mid p_{t-m} \right] - E_\infty \left[ \ln p \right] \right\} < \epsilon,
&\forall m \geq t_2.
\end{align*}
$$

We denote by $(\Omega, \mathcal{F}, \mathcal{P})$ the underlying probability space for the random variables $\{h_t\}$. Define $\{Y_t\}_{t \in \mathbb{N}}$ by $Y_t \equiv \xi_t - E(\xi_t) = \ln p_t - E[\ln p_t]$. We define the triangular array $\{X_{T,t} : t = 1, \ldots, T; \ T \in \mathbb{N}\}$ as $X_{T,t} \equiv \frac{t}{T} Y_t$, and the array of $\sigma$-fields $\{\mathcal{F}_{T,t} : t \in \mathbb{Z}; \ T \in \mathbb{N}\}$ as follows:

$$
\mathcal{F}_{T,t} = \begin{cases}
\{\emptyset, \Omega\}, & \text{if } t \leq 0, \\
\sigma(X_{T,1}, \ldots, X_{T,t}) & \text{if } 1 \leq t \leq T, \\
\mathcal{F}, & \text{if } t > T.
\end{cases}
$$

**Lemma A.3:** The triangular array $\{X_{T,t}, \mathcal{F}_{T,t}\}$ is a uniformly integrable $L^1$-mixingale.

**Proof of Lemma A.3:** Consider the non-trivial case where $T \in \mathbb{N}$, $1 \leq t \leq T$, and $1 \leq t - m \leq T$. Define $\{\psi_m\}_{m \geq 0}$ by

$$
\psi_m \equiv \sup_{t \geq m} \left( E \left\{ E_{t-m} \left[ \ln p_t \right] - E_\infty \left[ \ln p \right] \right\} + \left| E[\ln p_t] - E_\infty \left[ \ln p \right] \right| \right),
$$

where $E_{t-m}$ denotes the expectation with respect to the information at time $t - m$. By Lemma A.2, $\psi_m \to 0$. (See (17).) Note that
\[\|E(X_{T,t}|\mathcal{F}_T,(t-m))\|_1 \leq \psi_m, \quad \|X_{T,t} - E(X_{T,t}|\mathcal{F}_T,(t+m))\|_1 \leq \psi_{m+1},\]

where \(\| \cdot \|_1\) denotes the \(L^1\) norm. Q.E.D.

**Proof of Theorem 1:** By Lemma A.3 and Theorem 2(a) in Andrews (1988, p. 461),

\[\frac{1}{T} \sum_{t=1}^{T} X_{T,t} \to 0, \quad \text{in probability},\]

implying that:

\[\frac{1}{T^2} \sum_{t=1}^{T} t \xi_t \text{ goes to zero in probability.} \quad \text{Q.E.D.}\]

**Appendix B. Proof of Theorem 2**

For any given \(\delta > 0, \delta_\mu > 0\), and \(\mu = (\gamma_\mu, p_\mu^*) \neq (\gamma_0, p_0^*)\), let:

\[B((\gamma_0, p_0^*), \delta)^c = \{(\gamma, p^*) : \|\|\gamma, p^*) - (\gamma_0, p_0^*)\| \geq \delta\}, \quad \text{and:}\]

\[B(\mu, \delta_\mu) = \{(\gamma, p^*) \in B((\gamma_0, p_0^*), \delta)^c : \|\|\gamma, p^*) - (\gamma_\mu, p_\mu^*)\| < \delta_\mu\}.\]

Note that \(L_T((\gamma, p^*), \hat{\lambda}_T) = \frac{1}{T} \sum_{t=1}^{T} \left[u_t^2((\gamma, p^*), \hat{\lambda}_T) - u_t^2((\gamma_0, p_0^*), \hat{\lambda}_T)\right]\)

\[= \frac{1}{T} \sum_{t=1}^{T} \left[\left(2 \frac{P_{t+1}}{P_t} - 2 g_t((\gamma_0, p_0^*), \hat{\lambda}_T) + g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T)\right)\right.\]

\[\left.\times \left(g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T)\right)\right] = \]

\[\frac{2}{T} \sum_{t=1}^{T} u_t((\gamma_0, p_0^*), \hat{\lambda}_T) \left(g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T)\right)\]

\[+ \frac{1}{T} \sum_{t=1}^{T} \left(g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T)\right)^2.\]

Since the parameter space is compact, so is the complement \(B((\gamma_0, p_0^*), \delta)^c\). Therefore, it suffices to prove that the following inequality holds for any open ball \(B(\mu, \delta_\mu)\):
(18) \[ \liminf_{T \to \infty} \left\{ \inf_{(\gamma, p^*) \in B(\mu, \delta_u)} L_T((\gamma, p^*), \hat{\lambda}_T) \right\} \geq b > 0, \quad \text{in probability.} \]

To prove (18) it suffices to have both:

(19) \[ \sup_{(\gamma, p^*) \in B(\mu, \delta_u)} \frac{1}{T} \left| \sum_{t=1}^{T} u_t((\gamma_0, p_0^*), \hat{\lambda}_T) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \right| = o_p(1), \]

(20) \[ \liminf_{T \to \infty} \left\{ \inf_{(\gamma, p^*) \in B(\mu, \delta_u)} \frac{1}{T} \sum_{t=1}^{T} \left( g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T) \right)^2 \right\} \geq b > 0, \]

in probability. To prove (19), consider the following two lemmata:

**Lemma B.1:**

\[ \sup_{(\gamma, p^*) \in B(\mu, \delta_u)} \frac{1}{T} \left| \sum_{t=1}^{T} u_t((\gamma_0, p_0^*), \hat{\lambda}_T) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \right| - \sum_{t=1}^{T} \varepsilon_t \left| g_t((\gamma_0, p_0^*), \lambda_0) - g_t((\gamma, p^*), \lambda_0) \right| = o_p(1). \]

**Proof of Lemma B.1:**

\[ \frac{1}{T} \left| \sum_{t=1}^{T} u_t((\gamma_0, p_0^*), \hat{\lambda}_T) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) - \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma_0, p_0^*), \lambda_0) - g_t((\gamma, p^*), \lambda_0) \right) \right| \leq \frac{1}{T} \left| \sum_{t=1}^{T} \left( u_t((\gamma_0, p_0^*), \hat{\lambda}_T) - \varepsilon_t \right) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \right| + \frac{1}{T} \left| \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \right| + \frac{1}{T} \left| \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \lambda_0) \right) \right|. \]

Since \( EF^2(\omega) < \infty \), Breiman (1960) implies that there exists a constant \( K_2 \in \mathbb{R} \) such that:

(21) \[ \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \leq K_2 < \infty, \quad \forall \ T \in \mathbb{N}. \]

Let \( \overline{\gamma}, \overline{p^*} \) be finite upper bounds for \( \gamma \) and \( p^* \), respectively. Using the Cauchy-Schwarz’s inequality and (21), we conclude that, uniformly in \( (\gamma, p^*) \):

\[ \frac{1}{T} \left| \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \right| \]
\[ \leq \frac{1}{T} \sum_{t=1}^{T} |\varepsilon_t| \left| g_t((\gamma, p^*), \lambda_0) - g_t((\gamma, p^*), \hat{\lambda}_T) \right| \leq \]

\[ \leq \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \left[ g_t((\gamma, p^*), \lambda_0) - g_t((\gamma, p^*), \hat{\lambda}_T) \right]^2 \right)^{1/2} \]

\[ \leq K_2 \gamma \left( \frac{1}{T} \sum_{t=1}^{T} \left[ \min \left\{ \frac{p^*}{p_t}, 1 \right\} - \min \left\{ \left( \frac{\hat{\lambda}_T}{\lambda_0} \right)^T \frac{p^*}{p_t}, 1 \right\} \right] \right)^2 \]

\[ \leq K_2 \gamma \frac{p^*}{p} \left( 1 - \left( \frac{\hat{\lambda}_T}{\lambda_0} \right)^T \right) \to 0, \text{ in probability. Therefore:} \]

\[ \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \to 0, \text{ in probability,} \]

which is true in particular for $(\gamma, p^*) = (\gamma_0, p_0^*)$. Note that for any $(\gamma, p^*)$.

\[ \frac{1}{T} \sum_{t=1}^{T} \left( u_t((\gamma_0, p_0^*), \hat{\lambda}_T) - \varepsilon_t \right) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \leq \]

\[ \leq \frac{1}{T} \sum_{t=1}^{T} \left| u_t((\gamma_0, p_0^*), \hat{\lambda}_T) - \varepsilon_t \right| \left| g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right|. \]

Since \[ \left| u_t((\gamma_0, p_0^*), \hat{\lambda}_T) - \varepsilon_t \right| \leq \frac{\gamma_0 p_0^*}{p} \left( 1 - \left( \frac{\hat{\lambda}_T}{\lambda_0} \right)^T \right) \quad \text{and} \]

\[ \left| g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right| \leq 2\gamma, \]

we prove that, uniformly in $(\gamma, p^*)$,

\[ \frac{1}{T} \sum_{t=1}^{T} \left( u_t((\gamma_0, p_0^*), \hat{\lambda}_T) - \varepsilon_t \right) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \leq \]

\[ \leq 2\gamma \frac{\gamma_0 p_0^*}{p} \left( 1 - \left( \frac{\hat{\lambda}_T}{\lambda_0} \right)^T \right) \to 0, \text{ in probability. Q.E.D.} \]
Lemma B.2:

\[
\sup_{(\gamma, p^*) \in B(\mu, \delta_{\mu})} \frac{1}{T} \left| \sum_{t=1}^{T} u_t((\gamma_0, p_0^*), \hat{\lambda}_T) \left( g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \hat{\lambda}_T) \right) \right| = o_p(1)
\]

Proof of Lemma B.2: By Lemma B.1, it suffices to prove:

\[
(23) \sup_{(\gamma, p^*) \in B(\mu, \delta_{\mu})} \frac{1}{T} \left| \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma_0, p_0^*), \lambda_0) - g_t((\gamma, p^*), \lambda_0) \right) \right| = o(1), \hspace{1em} \text{a.s.}
\]

The law of large numbers of Breiman (1960) implies pointwise convergence, that is, for each given \((\gamma, p^*)\):

\[
\lim_{T \to \infty} \frac{1}{T} \left| \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma_0, p_0^*), \lambda_0) - g_t((\gamma, p^*), \lambda_0) \right) \right| = 0, \hspace{1em} \text{a.s.}
\]

To prove (23) it suffices to show that the family

\[
\left\{ \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma_0, p_0^*), \lambda_0) - g_t((\gamma, p^*), \lambda_0) \right) \right\}_{T \in \mathbb{N}}
\]

is equicontinuous.

Indeed, for any given \((\gamma_1, p_1^*), (\gamma_2, p_2^*) \in B(\mu, \delta_{\mu}),

\[
\left| \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \left( g_t((\gamma_1, p_1^*), \lambda_0) - g_t((\gamma_2, p_2^*), \lambda_0) \right) \right| \leq
\]

\[
\leq \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \left[ g_t((\gamma_1, p_1^*), \lambda_0) - g_t((\gamma_2, p_2^*), \lambda_0) \right]^2 \right)^{1/2}
\]

\[
\leq K_2 \left( \frac{1}{T} \sum_{t=1}^{T} \left[ g_t((\gamma_1, p_1^*), \lambda_0) - g_t((\gamma_2, p_2^*), \lambda_0) \right]^2 \right)^{1/2}. \hspace{1em} \text{Furthermore:}
\]

37
Lemma B.3: To prove (20) Q.E.D.

\[
\left| g_t((\gamma_1, p_1^*), \lambda_0) - g_t((\gamma_2, p_2^*), \lambda_0) \right| = \\
= \left| \gamma_1 \min \left\{ \frac{p_1^*}{p_t}, 1 \right\} - \gamma_2 \min \left\{ \frac{p_2^*}{p_t}, 1 \right\} \right| \\
\leq \left[ \gamma_1 \min \left\{ \frac{p_1^*}{p_t}, 1 \right\} - \gamma_2 \min \left\{ \frac{p_2^*}{p_t}, 1 \right\} \right] \\
+ \gamma_2 \left| \min \left\{ \frac{p_1^*}{p_t}, 1 \right\} - \min \left\{ \frac{p_2^*}{p_t}, 1 \right\} \right| \\
\leq \left[ |\gamma_1 - \gamma_2| + \gamma_2 \frac{|p_1^* - p_2^*|}{|p_t|} \right] \leq \left\{ 1 + \frac{\gamma_2}{p} \right\} \cdot ||(\gamma_1, p_1^*) - (\gamma_2, p_2^*)||. \]
Q.E.D.

To prove (20), consider the following two lemmata:

**Lemma B.3:** For each \((\gamma_\mu, p_\mu^*) \neq (\gamma_0, p_0^*)\), there exists a ball \(B(\mu, \delta_\mu)\) centered at \((\gamma_\mu, p_\mu^*)\) and a constant \(b > 0\), such that with probability one there exists \(T_3 \in \mathbb{N}\), \(T_3 = T_3(\{p_t\} \in \mathbb{N})\) with:

\[
\frac{1}{T} \inf_{(\gamma, p) \in B(\mu, \delta_\mu)} \sum_{t=1}^{T} \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right)^2 \geq b, \quad \forall \ T \geq T_3.
\]

**Proof of Lemma B.3:**

\[
\left| g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right| = \left| \gamma \min \left\{ \frac{p^*}{p_t}, 1 \right\} - \gamma_0 \min \left\{ \frac{p_0^*}{p_t}, 1 \right\} \right|.
\]

Case 1: \(\gamma_\mu = \gamma_0\). Since \((\gamma_\mu, p_\mu^*) \neq (\gamma_0, p_0^*)\), we conclude that \(p_\mu^* \neq p_0^*\). Without loss of generality we assume \(p_\mu^* < p_\mu^*\). There exists \(\varsigma > 0\) such that \(\alpha_1 = \mu_\infty([p_\mu^* + \varsigma, p_\mu^* + 2\varsigma]) > 0\). By the law of large numbers of Breiman (1960),

\[
\frac{\#\{1 \leq t \leq T : p_t \in [p_\mu^* + \varsigma, p_\mu^* + 2\varsigma]\}}{T} \to \alpha_1, \quad \text{with probability one.}
\]

Therefore, with probability one there exists \(T_1 \in \mathbb{N}\), \(T_1 = T_1(\{p_t\} \in \mathbb{N})\) with

\[
T \geq T_1 \Rightarrow \#\{1 \leq t \leq T : p_t \in [p_\mu^* + \varsigma, p_\mu^* + 2\varsigma]\} \geq \frac{\alpha_1}{2} T.
\]

Choosing \(\delta_\mu\) small enough such that \(p^* < p_\mu^* + \varsigma\) and \(\gamma p^* - \gamma_0 p_0^* \geq a > 0\) for some positive finite constant \(a > 0\), we conclude:

\((\gamma, p^*) \in B(\mu, \delta_\mu), \quad \text{and} \quad p_t \in [p_\mu^* + \varsigma, p_\mu^* + 2\varsigma] \Rightarrow\)
\[
\left| \gamma \min \left\{ \frac{p^*}{p_t}, 1 \right\} - \gamma_0 \min \left\{ \frac{p^*_0}{p_t}, 1 \right\} \right| = \frac{1}{p_t} |\gamma p^* - \gamma_0 p^*_0| \geq \frac{1}{p^*_\mu + 2\varsigma} |\gamma p^* - \gamma_0 p^*_0| \\
\geq \frac{a}{p^*_\mu + 2\varsigma}. \text{ Thus we have proved: } T \geq T_1 \Rightarrow \\
\frac{1}{T} \inf_{(\gamma, p^*) \in B(\mu, \delta_\mu)} \sum_{t=1}^{T} \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p^*_0), \lambda_0) \right)^2 \geq \frac{\alpha_1}{2} \left[ \frac{a}{p^*_\mu + 2\varsigma} \right]^2 > 0.
\]

Case 2: \( \gamma_\mu \neq \gamma_0 \). Without loss of generality we assume \( \gamma_\mu > \gamma_0 \). Take \( p^*_\mu > \mu \) and \( \varsigma > 0 \) such that: \( p^*_0, p^*_\mu > p^*_\mu - \varsigma > \mu \).

Let \( \alpha_2 \equiv \mu_\infty([\mu, p^*_\mu - \varsigma]) > 0 \). By the law of large numbers of Breiman (1960),
\[
\frac{\# \{1 \leq t \leq T : p_t \in [\mu, p^*_\mu - \varsigma] \}}{T} \to \alpha_2, \text{ with probability one.}
\]

Therefore, with probability one there exists \( T_2 \in \mathbb{N}, T_2 = T_2(\{p_t\}_{t \in \mathbb{N}}) \) with
\[
T \geq T_2 \Rightarrow \# \{1 \leq t \leq T : p_t \in [\mu, p^*_\mu - \varsigma] \} \geq \frac{\alpha_2}{2} T.
\]

Choosing \( \delta_\mu \) small enough such that \( p^* > p^*_\mu - \varsigma \), and \( \gamma - \gamma_0 \geq a > 0 \), for some positive finite constant \( a > 0 \), for all \( (\gamma, p^*) \in B(\mu, \delta_\mu) \), we conclude:

\[
(\gamma, p^*) \in B(\mu, \delta_\mu), \text{ and } p_t \in [\mu, p^*_\mu - \varsigma] \Rightarrow \frac{p^*}{p_t} > 1, \text{ and } \frac{p^*_0}{p_t} > 1 \Rightarrow \\
\Rightarrow |g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p^*_0), \lambda_0)| = \left| \gamma \min \left\{ \frac{p^*}{p_t}, 1 \right\} - \gamma_0 \min \left\{ \frac{p^*_0}{p_t}, 1 \right\} \right| = \\
= |\gamma - \gamma_0| \geq a > 0.
\]

We conclude that:
\[
T \geq T_2 \Rightarrow \frac{1}{T} \inf_{(\gamma, p^*) \in B(\mu, \delta_\mu)} \sum_{t=1}^{T} \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p^*_0), \lambda_0) \right)^2 \geq \frac{\alpha_1}{2} a^2 > 0.
\]

Q.E.D.
Lemma B.4: For all \((\gamma_\mu, p_\mu^*) \neq (\gamma_0, p_0^*)\), there exists a ball \(B(\mu, \delta_\mu)\) centered at \((\gamma_\mu, p_\mu^*)\), and a constant \(b > 0\), such that:

\[
\lim_{T \to \infty} \inf_{(\gamma, p^*) \in B(\mu, \delta_\mu)} \frac{1}{T} \sum_{t=1}^{T} \left( g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T) \right)^2 \geq b > 0,
\]

in probability.

Proof of Lemma B.4: By Lemma B.3, it suffices to show that:

\[
\sup_{(\gamma, p^*) \in B(\mu, \delta_\mu)} \frac{1}{T} \sum_{t=1}^{T} \left[ \left( g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T) \right)^2 - \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right)^2 \right] = o_p(1).
\]

Indeed:

\[
\begin{align*}
\left| \left( g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T) \right)^2 - \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right)^2 \right| &= \\
&= \left| g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T) + g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right| \times \\
&\quad \left| \left( g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \lambda_0) \right) - \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right) \right| \\
&\leq 4T \left[ \left| g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \lambda_0) \right| + \left| g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \lambda_0) \right| \right].
\end{align*}
\]

Therefore:

\[
\frac{1}{T} \sum_{t=1}^{T} \left[ \left( g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma_0, p_0^*), \hat{\lambda}_T) \right)^2 - \left( g_t((\gamma, p^*), \lambda_0) - g_t((\gamma_0, p_0^*), \lambda_0) \right)^2 \right] \\
\leq \frac{4T}{T} \sum_{t=1}^{T} \left[ \left| g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \lambda_0) \right| + \left| g_t((\gamma_0, p_0^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \lambda_0) \right| \right]
\]

40
and from the proof of (22) we know that:

$$\sup_{(\gamma, p^*)} \frac{1}{T} \sum_{t=1}^{T} \left| g_t((\gamma, p^*), \hat{\lambda}_T) - g_t((\gamma, p^*), \lambda_0) \right| \to 0, \text{ in probability,}$$

which is true in particular for $(\gamma, p) = (\gamma_0, p_0^*)$. Q.E.D.