Title
Essays on Estimation of a Nonlinear Commodity Price Model without a Closed-form Solution

Author
Zeng, Di

Publication Date
2012

Peer reviewed|Thesis/dissertation

Essays on Estimation of a Nonlinear Commodity Price Model without a Closed-form Solution

by

Di Zeng

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Brian D. Wright, Chair
Professor David Zilberman
Professor Michael Jansson

Fall 2012

Essays on Estimation of a Nonlinear Commodity Price Model without a Closed-form Solution

Abstract

Essays on Estimation of a Nonlinear Commodity Price Model without a Closed-form Solution

by

Di Zeng
Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Brian D. Wright, Chair

This thesis is about estimation of classic and modified versions of the rational expectations competitive storage model in the tradition of Gustafson (1958) (the storage model for short), an important economic theory of price determination of storable primary commodities.

The first chapter proposes and evaluates a procedure for approximating the optimal instruments under the context of the classic storage model. This procedure involves calibrating the unknown, true conditional variance function of price disturbance in the optimal instrument using the counterpart of an auxiliary storage model. Monte Carlo simulation suggests that this procedure brings small-sample efficiency gain relative to the benchmark Generalized Method of Moments (GMM) estimator of Deaton and Laroque (1992) and a few other alternatives at the sample size of 100. Its performance is also robust to parameterizations of the auxiliary model within moderate range from the true. This chapter also studies the estimators that require no preliminary estimation, which provide preliminary estimates for the proposed and other infeasible estimators. I find that a well-performing preliminary estimator does not contain instruments of lagged-more-than-one price or increasing transformations of lag prices, and does not use estimated optimal weighting matrix or adopt the continuous-updating approach of Hansen, Heaton and Yaron (1996). Therefore, an instrument of a constant plus reciprocal of lag one price and an identity weighting matrix in general form a good preliminary estimator.

Chapter 2 addresses two concerns about the usefulness of the theory of storage. While commodity speculators can induce serial dependence in price, Deaton and Laroque (1992, 1995 and 1996) argue that speculation explains only a small fraction of the observed autocorrelation in the actual data. Furthermore, the expected rate of return on storage implied by previous econometric estimates is implausibly small. This chapter addresses these two concerns about the validity of the theory of speculative storage by recognizing the downward trend in real price. The existence of a unique non-stationary equilibrium is proved for a rational-expectations competitive-storage model with a trend, and testable implications of the model are also derived. I show that, when a downward price trend in

part or all of the sample is ignored, the autocorrelation coefficient in price tends to be overestimated while the expected rate of return tends to be underestimated. Finally, I offer an empirical illustration of the trending storage model using annual corn price over the period from 1961 to 2005.

Chapter 3 discusses the empirical implications of the distributional misspecification of two nonlinear least squares estimators of a modified storage model with unbounded prices. The existence of infrequent, extremely low harvest generates extremely high cutoff price which is difficult to pass in finite periods. Meanwhile, due to the tiny chance of such events, it is easy for the practitioners to ignore them during the estimation and apply a false storage model with relatively low cutoff price. This chapter studies how such misspecification can affect the empirical implications of estimating the storage model. Surprisingly, I find that misspecified econometric models yield better estimates for the real interest rate; and the estimated cutoff price, actually captures the sharp turning point of the equilibrium price function. Therefore, though misspecified, the estimates are practically useful. Nevertheless, it is also emphasized that such interesting property of the two estimators should by no means be understood as a defense of ignoring the infrequent influential event in the asset pricing problems.

Mathematical proofs for general results and further discussions of a few econometric issues can be found in the Appendices. While a few theoretical results have been derived, this thesis relies heavily on Monte Carlo simulation and numerical functional approximation. Numerical methods turn out to be a convenient and many times necessary tools to study small-sample econometric problems when asymptotic results cannot provide an accurate approximation to the exact sampling.

For my beloved Bohan, my parents and my teachers

Contents

List of Figures

List of Tables

Acknowledgments

My graduate program requires each PhD student to write a second year econometric paper. Having passed quite a few ideas, I finally decided to replicate a well-known work and try my luck. This paper turned out to be Deaton and Laroque (1992), the path-breaking work that pioneered the formal empirical testing of the rational expectations competitive storage model in the tradition of Gustafson (1958) using the Generalized Method of Moments (GMM) of Hansen (1982).

This choice was not random. Professor Brian D. Wright originally from Australia has been my adviser since 2008. The economics of markets for storable commodities is his life-time interest and his book with Jeffery Williams on this topic is a must-read for those who have interests in the spot market. Their book relies heavily on numerical methods, the tools that even many today's young economists are not familiar with. I really want to thank Brian, who led me into this exciting field and guided me all the way. One step at a time is one of the important wisdoms I learnt from him, and always holding a positive attitude towards the difficulties in research is another.

Deaton and Laroque (1992) and a series of their work have posed serious challenges to the competitive storage model. Their arguments, however, could not convince me. The reason is simple, how can one be so confident about his empirical result if the sample size is only 88! So, I decided to argue that their empirical result is unreliable at small sample size.

I learnt a lot from reading Hansen, Heaton and Yaron (1996), a simulation work on the performance of various GMM estimators. Nevertheless, Hansen, Heaton and Yaron (1996) was not the key factor to the final success of my second year project. A former student of Brian, Carlo Cafiero from Italy, a dedicated young economist and skilled computer programmer, has offered great help during the process. He generously shared with me his estimation codes, which have facilitated all my research including this thesis.

My second year project did confirm the poor small-sample performance of the GMM estimator of Deaton and Laroque (1992) and found interesting pattern of bias that might be related to their later criticism of the storage model. Thanks to the kindness of the second year project committee, Professor Elisabeth Sadoulet, Professor Ethan Ligon and Professor Maxmilian Auffhammer, the paper was awarded the best project that year. The first chapter of this thesis is an extension to that project. It takes a more constructive attitude and tries to improve the small-sample performance of the GMM estimators of the competitive storage model rather than to criticize the existing ones in the literature.

The original purpose of Carlo to visit Berkeley was to work together with Brian and another of this former student, Eugenio Bobenrieth from Chile, who has already become a well-known economist. Juan, Eugenio's brother, a mathematician who started to gain interest in the "sub-field" of applied math, economics, is also part of their group. Eugenio gained interest in my GMM work when we came across each other on the third floor of Giannini Hall. He proposed opportunities for future collaborations. This is a great honor because they are all important figures in this field and of course I said yes.

Eugenio proposed quite a few projects that have attracted me for the rest of my PhD program here (and mostly likely more years in the future). One is the trending storage model, which is the focus of the second chapter of this thesis. Eugenio and Juan developed the original theory of storage behavior under trending prices. They were very generous to provide me with their unpublished manuscript which invokes a lot of my own thoughts, such as including population growth. According to the plan, we attempted to estimate the model using GMM. Unfortunately, our estimator does not satisfy the unusually assumed uniform convergence of the moment condition, a common condition for consistency of estimators. We, particularly Juan and I worked very hard on the proof for almost nine months. And eventually, Juan proved his superiority as a mathematician and produced a smart routine which can be intuitively summarized as a local uniform convergence on a shrinking area which always contains the estimates. The second chapter of this thesis, however, took a different and of course less exciting empirical approach.

The Wright group also proposed working on the so-called atom storage model which assumes a probability mass on the zero production. This mass changes the autoregression of price systematically. They worked intensively on the model. I occasionally offered some numerical help and most of time just pick up one or two pieces of their wisdom. While they are more interested in the properties of the model itself, I am more concerned about the misspecified estimation of the model: due to the fact that the chance of extreme events is tiny, it can easily be ignored by the applied econometricians at finite sample size even though the tiny chance can huge difference at far horizon. This motivates the study of Chapter 3 of this thesis. Interesting enough, I found that the misspecified estimators actually perform better than the correctly specified ones. This makes me strongly feel that, even after several years of studying limited-information estimators, as Brian asserted, we don't quite understand GMM.

They also proposed a few other projects that are less developed. Those projects include estimating the multi-storage model and maximum likelihood estimation of the atom model. Obviously, these will be the focus of our future work.

Outside the Wright group, Professor David Zilberman, my oral chair, has always been encouraging. We share the interest in in technology adoption and productivity. We had a paper together which becomes my first formal publication! We also have a working paper that I have delayed for too long and of course it will be guaranteed in the future.

Professor Gordon Rausser, another member of my oral committee and a true expert in the commodity market, has been encouraging me as well ever since my first class with him. Gordon is a great speaker. His style of presentation and his passion motivated me a lot.I learnt the importance of intuition for academic research in one his classes.

Professor Michael Jansson, a theoretical or true econometrician from the Economics Department at Berkeley, is my outside thesis member. He is so kind to continuously offer intellectual assistance to my applied work. It must be mentioned, that one of his suggestions to my oral presentation finally leads to the proposed method in the first Chapter of this thesis. This time, he offered another great suggestion. In fact, he tends to offer only one suggestion at at time and it is always a good one.

Professor Sofia Villas-Boas and Peter Berck offered me great help during the job market process. Since a revised version of the second Chapter is going to be used as a job market paper, their suggestions are also contributing to this thesis.

During the last few months of my PhD program, Yang Xie also from China and Ernesto Vellejos from Chile and student of Eugenio joined the group. They are energetic and eager to learn. I saw a better version of my former self in them. As Brian said, the younger has to be better than the older, otherwise there is no advancement. Of course, that path won't be easy at all, at least for me.

Chapter 1

Improving Small-sample Estimation of a Nonlinear Commodity Storage Model without Closed Solution

1.1 Introduction

The dramatic ascent of world food prices during 2007-08 and the ongoing “food vs. fuel” debate generated great interest in assessing the empirical relevance of the rational expectations competitive storage model in the tradition of Gustafson (1958). Empirical test of this important commodity pricing theory was pioneered by Deaton and Laroque (1992) using the Generalized Method of Moments (GMM). This empirical approach can estimate two important parameters, the real interest rate implied in the commodity market and a reference price level above which speculative storage is zero.

Despite its popularity among asset pricing problems, reliable GMM estimation of the storage model is challenging. First, the model does not yield a closed-form solution, so the parametrically testable structure is limited. Second, data are also limited: since high-frequency short-time-span data commonly seen in empirical finance contains much noise and does not include enough influential commodity booms, this literature usually uses low-frequency long-time-span data such as annual average prices. Also, since structural changes in commodity markets are common, time relevant inference further restricts sample size. One consequence of these empirical difficulties, as asserted by Deaton and Laroque (1992), is that the “discount rates are notoriously hard to estimate”.

In the context of the storage model, I propose a procedure for approximating the infeasible optimal instrument with which a GMM estimator attains the asymptotic Cramér-Rao bound for the given conditional restriction. This procedure involves calibrating the unknown, true conditional variance function of disturbance in the optimal instrument using the numerically represented counterpart of an auxiliary storage model. The rescaling is conducted by first estimating a reference point on the true conditional variance function and then rescaling the auxiliary conditional variance function to align its counterpart with the estimated one. In principle, such rescaling can provide a reasonable approximation to the true function even when the chosen auxiliary model is quite different from the true. The most advantage of this procedure is that the important features of the true function that cannot be evaluated directly are largely maintained, if not exactly, without being affected too much by the small-sample limitation. Using numerical experiments, I show that this procedure is small-sample efficient relative to the benchmark estimator of Deaton and Laroque (1992) and a few other alternatives; performance of the result instrument is also robust to parameterizations of the auxiliary model within moderate range from the true. This chapter also investigates the small-sample performance of basic instruments that contain no unknown part. I find: 1) iterative use of the estimated optimal weighting matrix undermines small-sample performance and so does the continuous updating approach; 2) instruments formed by lag one prices perform better than those including also older prices; and 3) decreasing functions of lag one price perform better than increasing ones, particularly the commonly used lag-one price itself in the literature.

The proposed procedure constructs an approximant to the optimal instrument estimator given preliminary estimates. The study of the basic instruments attempts to locate a reliable simple estimator that requires no preliminary step. The two merge when the latter feeds preliminary estimates to the former. As a result, this chapter offers a complete picture of the proposed method as well as a reliable simple estimator if convenience of application is strongly preferred.

Other than its practical usefulness, this chapter contributes also from a few methodological perspectives. First, it provides insights to indirectly implement the infeasible optimal instruments, use of which is complicated by the fact that the conditional expectation within does not have known functional form. Nonparametric methods provide a way to circumvent this difficulty but can become unreliable when sample size is seriously limited (Anatolyev 2007). The proposed approach, though in the end introduces inefficiency, is justified if the data situation is not going to change soon. It can provide more stable approximations, with much less estimation effort, over time series filtrations, than those from approaches which result in instruments that can be seriously sample-dependent at small sample size.

Second, this chapter offers an alternative to the simulation-based methods to improve the finite-sample properties of the GMM estimators. The need for knowing the exact functional form of conditional restrictions limits the applicability of GMM estimators to asset pricing problems. As a result, the simulation-based methods of moments, which numerically represent the moments that cannot be evaluated directly using an auxiliary model, have been developed and proposed. Michaelides and Ng (2000), however, found that the performance of such methods in estimating the storage model is sensitive to choice of auxiliary models and inferior to the Pseudo Maximum Likelihood (PML) estimator. Cafiero, et al. (2010) found that PML does not perform as well as the Deaton and Laroque's (1992) GMM estimator in inferring the reference price level when the real interest rate is fixed at its true value. This chapter brings finite-sample efficiency gain to Deaton and Laroque's estimator with robustness to specifications of the auxiliary model. Moreover, unlike the simulation-based methods, since the proposed procedure works only with instruments, misspecification of the conditional variance function cannot alter consistency.

Third, this chapter adds evidence to the literature on finite-sample behavior of the GMM estimators. For example, while I confirm poor small-sample performance of the estimated optimal weighting matrix in line with the finding of Hansen Heaton and Yaron (1996), I also find that the continuous-updating estimator they propose does not work well. Moreover, my examination of lag-length provides evidence of small-sample efficiency loss caused by using lag price beyond the degree of freedom, to the contrary of what asymptotic theory would suggest. Finally, I relate the observed advantage of instruments with decreasing functions of lag one price over other transformations to the instruments' implicit effect of weighting residuals and the fact that the conditional variance function is weakly increasing in lag one price in storage model.

The rest of the chapter is organized as follows. Section 2 reviews the storage model. Section 3 introduces the estimator and the approximating method. Section 4 evaluates the proposed method and section 5 studies basic instruments. Section 6 concludes. Mathematical proofs and some extra simulations can be found in the appendix.

1.2 Review of the Storage Model

Consider a competitive market for a single storable commodity. Time is discrete. The real price of the commodity at period t is p_t. There are two types of agents, consumers whose excess demand for the commodity depends only on its current price and inventory holders who store the commodity from one period to the next. Both types of agents have rational expectations.

The exogenous, stochastic excess supply at period t is ω_t. The non-stochastic, continuous, strictly decreasing inverse consumption demand function is $f(\cdot)$. The consumption at period t is therefore $c_t = f^{-1}(p_t)$.

Inventory holders have access to a simple constant returns storage technology: one unit of commodity stored at t yields $(1-\rho)$ units at $t+1$, where ρ is the deterioration rate. Let x_t denote the inventory at period t. Then the availability at period t is $z_t = c_t + x_t$ and the equation of motion is $z_{t+1} = (1-\rho)\,x_t + \omega_{t+1}$. An important feature of the storage model is the emphasis on the constraint that market as a whole cannot hold negative physical inventories, i.e. $x_t \geq 0,\ \forall t \geq 0$.

Assuming that inventory holders are risk neutral and competitive, and have access to a perfect capital market where the rate of interest is r, the planner's problem corresponding to the decentralized economy is to maximize the sum of discounted future surplus streams with respect to x_t:

$$V_t(z_t) = \int_0^{z_t - x_t} f(q)\,dq + \sum_{j=t+1}^{\infty} \frac{1}{(1+r)^{j-t}} E_t\left[\int_0^{z_{t+1}-x_{t+1}} f(q)\,dq\right] \tag{1.1}$$

$$s.t.\ x_t \geq 0 \text{ and } z_{t+1} = (1-\rho)\,x_t + \omega_{t+1},$$

where $E_t[\cdot]$ is the conditional expectation on the available information at period t.

The Euler equation is

$$x_t > 0 \text{ if } 0 = \frac{\partial V(z_t)}{\partial x_t} = -f(x_t) + \frac{1-\rho}{1+r} E_t f(c_{t+1})\,; \tag{1.2}$$

$$x_t = 0 \text{ if } 0 \geq \frac{\partial V(z_t)}{\partial x_t} = -f(x_t) + \frac{1-\rho}{1+r} E_t f(c_{t+1})\,. \tag{1.3}$$

Denoting $\gamma \equiv (1+r)/(1-\rho)$, rearranging of the equation above yields:

$$x_t > 0 \text{ if } f(c_t) = \frac{1}{\gamma} E_t f(c_{t+1})\,;\ x_t = 0 \text{ if } f(c_t) \geq \frac{1}{\gamma} E_t f(c_{t+1})\,. \tag{1.4}$$

Since $f(\cdot)$ is strictly decreasing, the Euler equation can be compactly written as:

$$p_t = f(c_t) = \max\left\{ f(z_t)\,, \frac{1}{\gamma} E_t f(c_{t+1}) \right\}, \tag{1.5}$$

$$s.t.\ z_{t+1} = (1-\rho)\left(z_t - f^{-1}(p_t)\right) + \omega_{t+1}.$$

This Euler condition says if carrying forward one unit generates profits in expectation, inventory holders will store until the current price equals the expected forward price after due allowance for carrying cost; If holding inventories generates expected loss to begin with, nothing will be stored and all available supply will be consumed.

Suppose the equilibrium at period $t+1$ is a function $p_{t+1}(\cdot)$ of the state variable z_{t+1}. Then, $p_t(\cdot)$, the equilibrium at period t satisfies:

$$p_t(z) = \max\left\{f(z), \frac{1}{\gamma}E_t p_{t+1}\left\{\omega_{t+1} + (1-\rho)\left(z - f^{-1}(p_t(z))\right)\right\}\right\}. \tag{1.6}$$

Denote $p_t^* \equiv \gamma^{-1}E_t p_{t+1}(\omega_{t+1})$. Since $p_t(\cdot)$ and $p_{t+1}(\cdot)$ are weakly decreasing, we have the following relation:

$$E_t p_{t+1} = \gamma p_t, \text{ if } p_t \leq p_t^*; E_t p_{t+1} = \gamma p^*, \text{ if } p_t > p_t^*. \tag{1.7}$$

Under certain regularity conditions, there exists a unique stationary rational expectations equilibrium. That is, $p_t(z) = p_{t+1}(z),\ \forall t \geq 0$ and $\forall z \geq 0$. Then, assuming *i.i.d.* ω_t, we must have $p_t^* \equiv p^*,\ \forall t \geq 0$. Thus, we have the following auto-regressive relation:

$$u_t = p_t - \gamma \min\{p^*, p_{t-1}\}, \quad t \in \mathbb{N}. \tag{1.8}$$

I numerically solve a storage model backwards under: $f(c) = 600 - 5c$, $r = 5\%$, $\rho = 0$, ω_t *i.i.d.* standard normal with mean 100 and standard deviation 10 (see Figure 1.1). The implied cutoff price is $p^* = 114.1243$, depicted as the kink on the price function.

Let $\zeta(p_t)$ denote the function of variance of p_{t+1} in terms of p_t:

$$\zeta(p_t) \equiv Var\left[p\left((1-\rho)s(p_t) + \omega_{t+1}\right)\right], \tag{1.9}$$

where $p(\cdot)$ is the equilibrium price function, $s(\cdot)$ the optimal storage rule and ω_{t+1} the random harvest shock at $t+1$. Specifically, $s(p_t) \equiv p^{-1}(p_t) - f^{-1}(p_t)$.

Under the parameterization above, I numerically solve for the conditional distribution of forward price given current price (see Figure 1.2). We can easily see that higher current price leads to larger variation at higher price level. We also see that when the current prices are higher than the cutoff price (120 and 150 in the figure), the conditional distributions of forward prices are identical because carry-in stocks are identical at zero for periods following stock-out observations. This figure is a very clear presentation of conditional heteroskedasticity we face in the estimation of the storage model.

I also plot the conditional variance function (see Figure 3). This function is continuous, strictly increasing and strictly convex below the cutoff price and flat beyond it, consistent with the observation in Figure 1.2. Knowing the conditional variance function can greatly facilitate estimation. For example, it gives the weight required in the weighted least squares. Unfortunately, as we have seen, this function can only be represented numerically and unknown in practice.

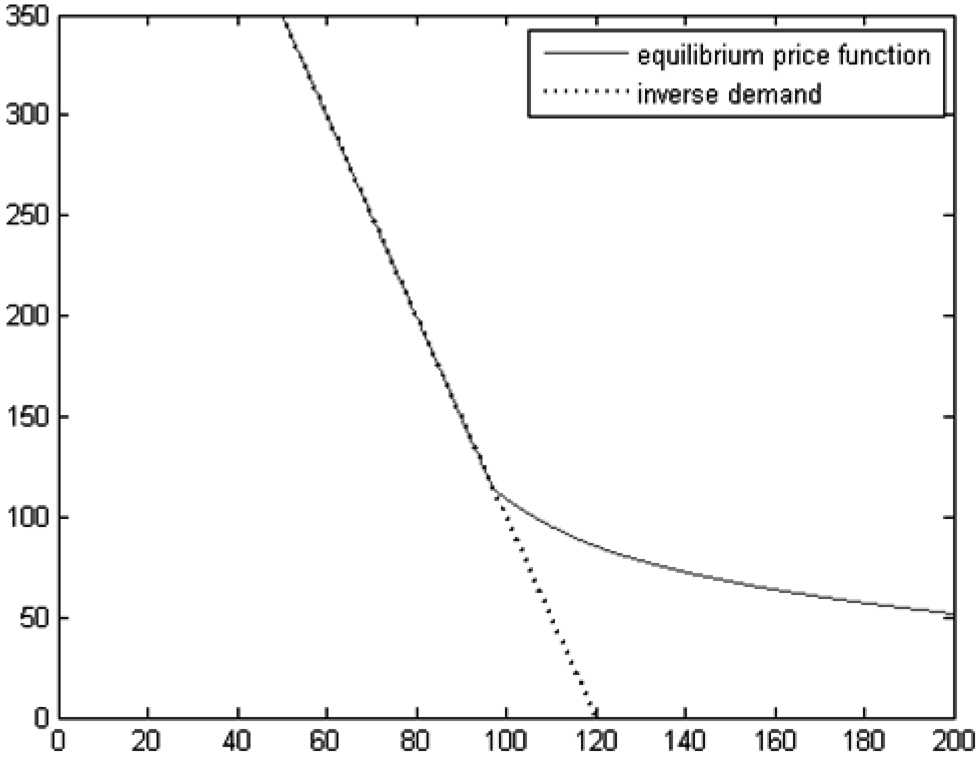

Figure 1.1: Equilibrium price function

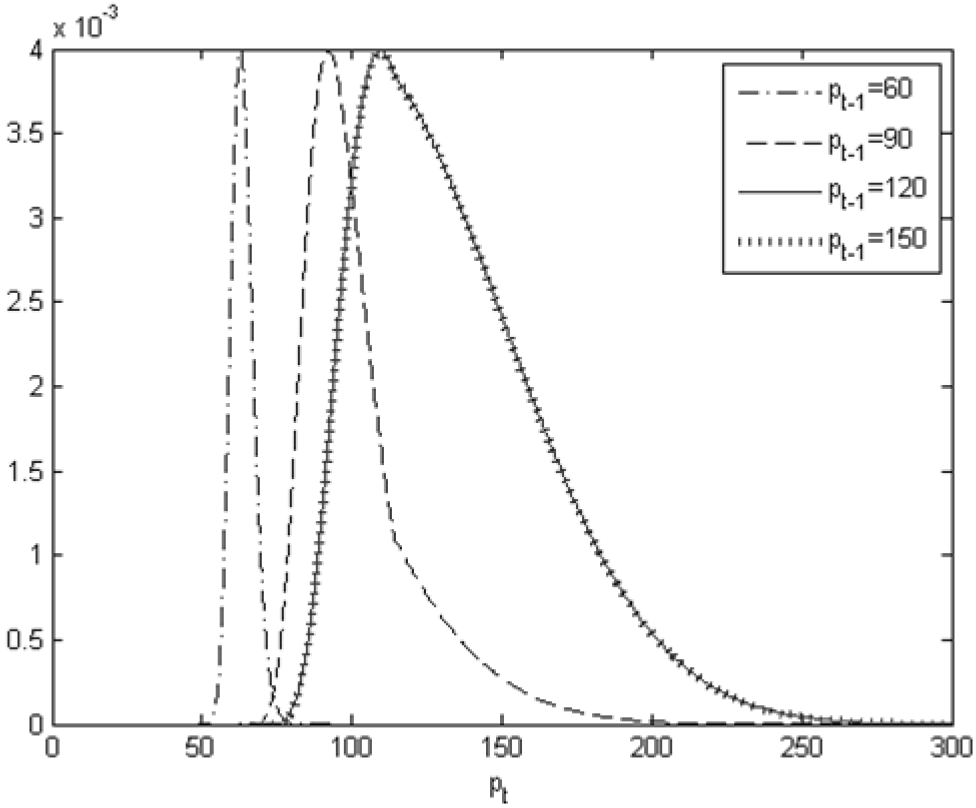

Figure 1.2: Conditional price distributions

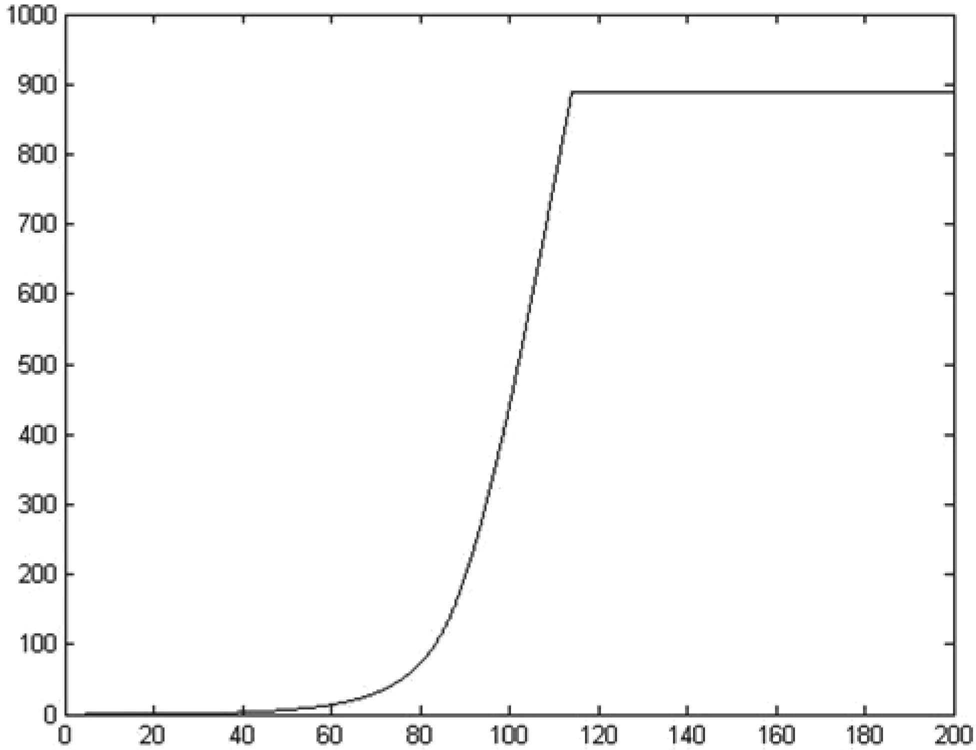

Figure 1.3: Conditional variance function of price disturbance

1.3 The GMM Estimators of the Storage Model

Discrete-time models of the optimizing behavior of economic agents often lead to first-order conditions of the form:

$$E_{t-1}u_t(\theta_0) = 0, \qquad t \in \mathbb{N}, \tag{1.10}$$

where θ_0 is the true parameter.

In the case of the storage model, $\theta_0 = (\gamma, p^*)$ and

$$u_t(\theta_0) = p_t - \gamma \min\{p^*, p_{t-1}\}, \qquad t \in \mathbb{N}. \tag{1.11}$$

Let z_t denote a $q \times 1$ vector of variables with finite second moments. The variablesz_t are in the agent's information set Ω_{t-1} and observed by the econometricians. Define the process $\phi_t(\theta_0)$ by:

$$\phi_t(\theta_0) = u_t(\theta_0) \otimes z_t, \qquad t \in \mathbb{N}, \tag{1.12}$$

where $\otimes$ is the Kronecker product. Then, we must have:

$$E_{t-1}\phi_t(\theta_0) = 0, \qquad t \in \mathbb{N}. \tag{1.13}$$

By the law of iterated expectation, we have the unconditional moment conditions:

$$E\phi_t(\theta_0) = 0, \qquad t \in \mathbb{N}. \tag{1.14}$$

Let

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \phi_t(\theta), \tag{1.15}$$

where $\theta \in \Theta$, the parameter space.

Then, θ_0 can be estimated by choosing $\theta \in \Theta$ to minimize the objective function given by

$$g_T(\theta)' W_T g_T(\theta), \tag{1.16}$$

where W_T is a positive semi-definite matrix that converges to a constant positive definite matrix. For given unconditional moment conditions, asymptotic efficiency bound is attained if optimal weighting matrix is used. In the current context, a consistent estimator for the optimal weighting matrix is:

$$\left(Z_T' \hat{D}_T Z_T\right)^{-1} \tag{1.17}$$

where

$$Z_T = [z_1, ..., z_T], \quad \hat{D}_T = Diag\left\{\hat{u}_1^2, ..., \hat{u}_T^2\right\},$$

and $\hat{u}_t$ is residual from previous consistent estimation. Iteration using previous consistent estimates is one way to apply the optimal weighting matrix. Another is to estimate it simultaneously with the sample moments, as in the continuous updating approach of Hansen, Heaton and Yaron (1996).

For given conditional moment conditions, the optimal instrument gives rise to the asymptotically efficient estimator whose asymptotic variance cannot be reduced by adding additional instruments.[1] For no-arbitrage condition (or Euler equation) of stochastic dynamic rational expectations models, if the innovation can be written in the form of: $u_t(\theta_0) = y_t - f(x_t, \theta_0)$, the optimal instrument takes the form:[2]

$$\frac{\partial f(x_t, \theta_0)/\partial \theta'}{E_{t-1}\left[u_t^2(\theta_0)\right]}. \tag{1.18}$$

For the storage model, the optimal instrument is:

$$\left[\frac{\min\{p_0^*, p_{t-1}\}}{E\left[u_t^2(\theta_0)\,|p_{t-1}\right]}, \frac{\gamma_0 I_{\{p_0^* \le p_{t-1}\}}}{E\left[u_t^2(\theta_0)\,|p_{t-1}\right]}\right]. \tag{1.19}$$

Figure 1.4 plots the numerically solved optimal instrument under the same numerical parameterization as before. Theory predicts that most prices will be below the cutoff price. Thus, the segment of the optimal instrument that will be applied most is the part below the cutoff price.

Application of the optimal instrument is complicated by its requirement of knowing the exact functional form of the conditional variance. For models without closed solution, this is difficult to satisfy.

The unknown parameters in the optimal instrument can either be replaced by preliminary estimates in the plug-in sense or estimated simultaneously with the sample moments in the continuous-updating sense.

The conditional expectations can be estimated either non-parametrically or parametrically by imposing an auxiliary parameterization. When data are abundant non-parametric approaches are reasonable choices. In contrast, while parametric estimation by imposing an auxiliary parameterization will in the end introduce inefficiency, the basic insight of this approach is justified if the data situation is not going to change soon. The difficulty, however, lies at the choice of an appropriate auxiliary parameterization.

Naturally, an appropriate parameterization can come from a storage model with a parameterization that is most likely different from the true model. I propose to use the calibrated conditional variance function of an auxiliary storage model in place of the true conditional variance function in the optimal instrument. Specifically, denote the

[1] Amemiya (1977) provided the optimal instrument for a nonlinear simultaneous equation model with homoscedastic and serially uncorrelated errors. Hansen (1985) derived the optimal instruments for a conditional mean model with dependent observations. Chamberlain (1987) found that the optimal instrument estimator attains the semi-parametric efficiency bound for conditional moment restrictions.

[2] See Hansen (1985), Hansen, Heaton and Ogaki (1988). A nice survey can be found in Anatolyev (2007).

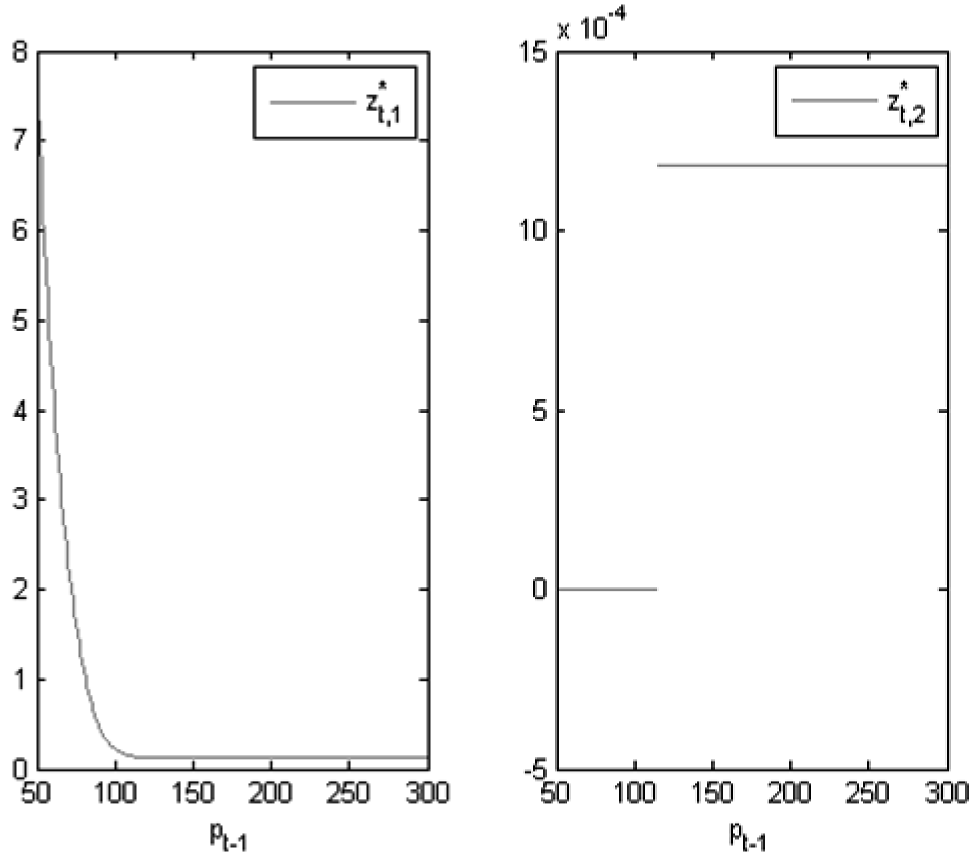

Figure 1.4: Optimal instrument to estimate storage model

true model as Model I and the auxiliary model for approximating Model I's conditional variance function as Model II. Let $\zeta_{II}(\cdot)$ denote the numerically represented conditional variance function of Model II. I propose to approximate the conditional variance function of Model I using:

$$\left(\frac{p_I^*}{p_{II}^*}\right)^2 \cdot \zeta_{II}\left(p_{t-1} \cdot \frac{p_{II}^*}{p_I^*}\right), \tag{1.20}$$

where p_t is the observable prices of Model I and, p_I^* and p_{II}^* are the cutoff prices of Model I and II, respectively. Note, however, that:

Lemma 1. *For given conditional moment conditions, a GMM estimator with instrument z is equivalent to the GMM estimator with instrument $c \otimes z$ for any non-zero constant c.*

Proof. Trivial. □

Then, it is equivalent to replace the conditional variance function of Model I by

$$\hat{\zeta}_I\left(p_{t-1}; p_I^*, p_{II}^*\right) = \zeta_I\left(p_{t-1} \cdot \frac{p_{II}^*}{p_I^*}\right). \tag{1.21}$$

Thus, the instrument for estimating Model I implied by the approximating method above is

$$\left[\frac{\min\left\{p_I^*, p_{t-1}\right\}}{\hat{\zeta}_I\left(p_{t-1}; p_I^*, p_{II}^*\right)}, \frac{\gamma_I I_{\left\{p_I^* \leq p_{t-1}\right\}}}{\hat{\zeta}_I\left(p_{t-1}; p_I^*, p_{II}^*\right)}\right], \tag{1.22}$$

where p_{II}^* is known once Model II is chosen, and p_I^* and γ_I can be replaced by previous consistent estimates.

The asymptotics of the resulting estimator is standard.

1.3.1 The Robustness of the Approximating Method

While in practice the auxiliary model (i.e., Model II) is chosen with necessary discrepancy from the unknown true model (i.e., Model I), various parameterizations may actually share the same (up to a scalar) or similar conditional variance function. In the following analysis, I first rely on the theory of under-identification to show the equivalence of the conditional variance function of Model II to those implied by a range of parameterizations. I then numerically compare $\hat{\zeta}_I(p_{t-1})$ with the true conditional variance function when the auxiliary model is outside the equivalence set.

Let x_i, f_i, $p_i(\cdot)$ and p_i, $i = \mathbb{I}, \mathbb{II}$ denote the quantity variable, the inverse demand function, the equilibrium price function and price of Model i. We have the following equivalence result:

Proposition 2. *If Γ is an additive monotonic transformation on the real line, $x_{II} = \Gamma x_I$ and $f_{II}(x_{II}) = \lambda f_I(x_I)$, $\lambda > 0$, then $\lambda p_I(x_I) = p_{II}(x_{II})$ and $\zeta_{II}(p_{II}) = \lambda^2 \zeta_I(p_I)$.*

Proof. See Appendix A. □

Examples of the proposition include:

- Changing the price units: $x_I = x_{II}$ and $\lambda > 0$.
- Shifting and rescaling quantities: Suppose Model I and II are identical except for quantity units. They have inverse demand function $f(c) = (Ac + B)^\eta$, where $\eta < 0$. Then, if $x_{II} = x_I/\sigma + (\alpha^{-1} - 1) \cdot B/A$, by Proposition 2, $\alpha^\eta p_I(x_I) = p_{II}(x_{II})$ and $\zeta_{II}(p_{II}) = \alpha^{2\eta}\zeta_I(p_I)$.

Therefore, for Model I and II satisfying the conditions in Proposition 1, the rescaled conditional variance function of Model II and the conditional variance function of Model I are equivalent in GMM estimation.

Of course, there is no reason to believe that Model I and II will fall into the equivalence set as defined in Proposition 2. It is necessary to numerically study the quality of the approximating method for Model I and II that fall outside the equivalence set. Let Model I have the same parameterization as in previous numerical examples so we can focus on the parameterizations of Model II in Table 1.1.

Table 1.1: Perturbation of Parameterizations of the Auxiliary Model

	a	b	r	μ	σ	p^*
			Model I			
0	600	5	0.05	100	10	114.1243
			Model II			
1	600	4.5	0.05	100	10	156.9690
2	600	5.5	0.05	100	10	72.5473
3	600	5	0.04	100	10	116.4020
4	600	5	0.06	100	10	112.0491
5	600	5	0.05	100	8	109.2842
6	600	5	0.05	100	12	119.1011

Note. p^* is calculated after solving the model under the corresponding parameterization.

I present the rescaled conditional variance function of Model II together with that of Model I. For ease of comparison, I align their kinks.

From Figure 1.5, we see that the rescaled conditional variance function is more sensitive to changes in b: 10% perturbation in b generates bigger swing of the rescaled conditional variance function than 20% perturbation in other parameters. Even so, in general, we do not see any disturbing change in the rescaled conditional variance function following the perturbations in the covered range.

It is important, however, to clarify that the above comparison assumes knowing the kink of the conditional variance function of Model I, which, in practice, has to be estimated. The estimated kink can be inaccurate at small sample size, and thus the rescaled

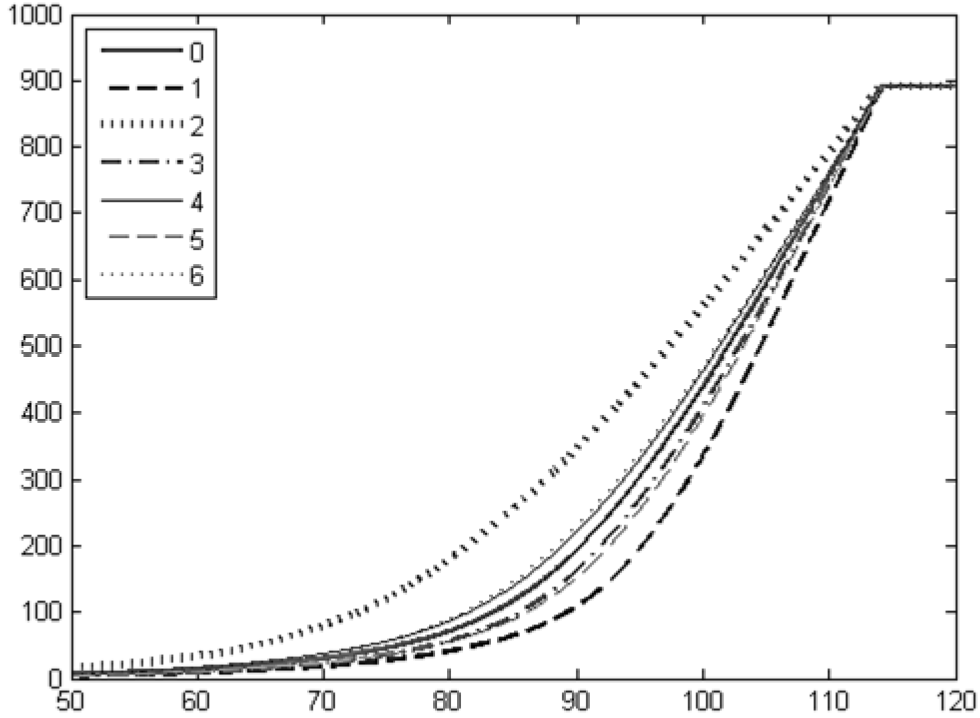

Figure 1.5: Aligned conditional variance functions under various parameterizations

conditional variance function of Model II can be more different from the true than in Figure 1.5. Therefore, the actual performance of the proposed method must be tested using Monte Carlo simulation.

1.4 The Performance of the Approximate Optimal Instrument

I denote as OIV the true optimal instrument estimator (known in simulation), DL1 the Deaton and Laroque (1992) estimator, DL2 a two-step GMM estimator uses DL1 as first step and estimated optimal weighting matrix for second step, BIV an estimator that uses instrument $\left[1, p_{t-1}^{-1}\right]$ and identity weighting matrix, and AOIVi, i=0,1,...,6, the GMM estimators using the proposed approximate optimal instrument whose functional form is from Model i in Table 1.1. The unknown parameter in the approximate optimal instrument is replaced by BIV. I will argue in the next section why BIV is chosen as the preliminary estimator.

In this and all the later simulations, the parameterization is the same as the numerical examples in the last section unless specially notified. The total number of replications for each experiment is 10000 and sample size for each replication is always 100, small but about the size of reliable world annual average commodity prices. In each table, 25%, 50% and 75% stand for the 25, 50 and 75 percentile values of each collection of estimates, STD stands for the standard deviation of each collection and RMSE, the root mean square error, is a statistic for the average deviations of the estimates from the true. RMSE will be the major criterion for evaluating estimators, while sometimes I pay closer attention to percentile values.

By comparing OIV with others, we can obtain some idea about how big the gap is between asymptotic efficiency and the small-sample practice. By comparing DL1 with AOIVi, we know if the proposed method is after all useful. By comparing among AOIVi, we know how sensitive the performance is to the misspecification of the conditional variance function. The simulation result is reported in Table 1.2.

First of all, the small-sample performance of the true optimal instrument is not the best for either γ or p^*: all AOIVi are better in estimating both parameters and all other estimators except DL2 are better in estimating p^* under the RMSE criterion.

Tauchen (1996) found that just-identified estimator with true optimal instruments exhibits smaller mean bias than the over-identified two-step estimator with optimal weighting matrix, but produces larger standard deviation. The finding here is the opposite for estimating γ, while for estimating p^*, OIV produces smaller mean bias and standard deviation than DL2. Given that his work is under a different context and his sample size is 50 and 75 while mine 100, it is difficult to extract deep insight from this comparison. But since the optimal instrument is efficient asymptotically by definition, sample size should

Table 1.2: The performance of approximate optimal instrument

			$\gamma = 1.05$			
	Mean	25%	50%	75%	STD	RMSE
OIV	1.0543	1.0422	1.052	1.0635	0.0169	0.0174
DL1	1.0551	1.0397	1.0522	1.0675	0.0216	0.0222
DL2	1.0532	1.0383	1.0502	1.0651	0.0209	0.0211
BIV	1.0542	1.0408	1.0518	1.0649	0.0185	0.0190
AOIV0	1.0537	1.0419	1.0517	1.0629	0.0164	0.0168
AOIV1	1.0540	1.0420	1.0518	1.0633	0.0166	0.0170
AOIV2	1.0532	1.0409	1.0511	1.0629	0.0170	0.0173
AOIV3	1.0539	1.0421	1.0518	1.0632	0.0164	0.0168
AOIV4	1.0536	1.0418	1.0515	1.0628	0.0164	0.0168
AOIV5	1.0538	1.0420	1.0517	1.0631	0.0164	0.0169
AOIV6	1.0536	1.0418	1.0515	1.0629	0.0164	0.0168
			$p^* = 114.124$			
	Mean	25%	50%	75%	STD	RMSE
OIV	112.797	107.249	112.529	117.964	8.658	8.759
DL1	112.218	106.124	111.401	117.484	8.544	8.753
DL2	110.475	104.517	109.651	115.589	8.447	9.202
BIV	112.316	106.687	111.689	117.405	7.934	8.137
AOIV0	113.095	107.336	112.558	118.137	8.395	8.458
AOIV1	113.052	107.297	112.529	118.110	8.398	8.466
AOIV2	113.202	107.385	112.559	118.290	8.464	8.514
AOIV3	113.064	107.316	112.559	118.118	8.388	8.454
AOIV4	113.120	107.336	112.573	118.153	8.405	8.464
AOIV5	113.072	107.329	112.556	118.110	8.394	8.460
AOIV6	113.118	107.346	112.570	118.167	8.401	8.461

Note. 10000 replications with sample size 100.

For AOIVi, the first step instrument is always $\left[1, p_t^{-1}\right]$.

matter for its actual performance.[3]

Second, all approximate optimal instrument estimators work satisfactorily. They produce smaller RMSE than all other estimators, except that BIV yields smaller RMSE for p^*.

The relative performance of AOIVi and BIV, however, makes comparison between them difficult. On one hand, BIV produces more reliable estimates of p^* under the RMSE criterion. This advantage justifies the use of BIV as the preliminary estimator: it provides accurate estimate of the kink and therefore better rescaling of the auxiliary conditional variance function. On the other hand, AOIVi enjoys great advantage in estimating γ over all others including BIV: the RMSE of AOIV0 is 24% smaller than that of DL1 and 11.6% smaller than that of BIV.

A closer scrutiny reveals that AOIVi produces closer mean, 25 percentile and median to the true cutoff price than BIV. But, it tends to have a thicker right tail as evident by its further 75 percentile value, and bigger standard deviation and RMSE. Therefore, if outlier estimates of p^* are abandoned or an upper bound for it is imposed in estimation, AOIVi may outperform BIV in estimating p^*.

Given that AOIVi is designed to approximate OIV, it is surprising to see that AOIVi performs better than its true counterpart. This is particularly interesting when compared with Angrist and Krueger (1995), Altonji and Segal (1996), and Angrist, Imbens and Krueger (1999), who blamed the correlation between the estimates from the first stage and error terms of the second stage for small-sample bias. My simulation suggests that such correlation may not always hurt second-step estimation.

Equally important is that the performance of AOIVi is not very sensitive to the misspecification of the conditional variance function. The range of variation in RMSE is within 0.0005 for γ and 0.06 for p^*, or about 3% and 0.7% of their smallest RMSEs, respectively. In contrast, the difference in RMSE between DL1 and DL2 are easily over 0.01 for γ and 0.4 for p^*. This result should not be surprising as we have seen above that the conditional variance functions of disturbances under various parameterizations do not disturbingly different when their kinks are aligned. Given the satisfactory performances of all AOIVi, such robustness to specifications of auxiliary model obviously adds credits to the proposed method.

Nevertheless, it is still important to be aware that the variation in b of the inverse demand function, $f(c) = a - b \cdot c$, tends to affect the estimation more than equal percentage changes in other parameters. This is consistent with the observation above that change in

[3]That the asymptotic optimal instrument is not performing well at small-sample size is not difficult to understand. We know that the covariance matrix of θ_T satisfies

$$Cov(\theta_T) \geq \frac{\partial E[\theta_T]}{\partial \theta} [I(\theta)]^{-1} \left(\frac{\partial E[\theta_T]}{\partial \theta} \right)',$$

where $I(\theta_0)$ is the Fisher information matrix. If $E(\theta_T) \equiv \theta$, the Cramér-Rao bound reduces to $Cov(\theta_T) \geq [I(\theta)]^{-1}$, and is attained asymptotically by a consistent estimator with optimal instrument. At finite sample $\partial E[\theta_T] / \partial \theta$ is not always identity and thus the asymptotic optimal instrument is not efficient.

b tends to swing the rescaled conditional variance function the most. How much this can hurt the applicability of the method is difficult to measure without specifying a procedure for choosing the auxiliary model. If the axillary model is implied by ML estimates using the same data, then the effect might be limited because change in b *ceteris paribus* should noticeably alters the price dynamics (evident by the relatively big consequent change in the conditional variance function) and this should restrain ML estimate of b from going too far from the true.[4]

Also worth mentioning is the choice of the preliminary estimator for the proposed method. As the preliminary estimator for AOIVi, BIV is already better than the Deaton and Laroque (1992) estimator. If the practitioners prefer a simple estimator, then BIV can be a good choice. A natural question then, either for a complete evaluation of the proposed method or for the sake of finding a simple good estimator, is what properties a good basic instrument, that does not contain unknown part and is ready to use, should have. This will be answered seriously in the next section.

1.5 The Performance of the Basic Instrument

Deaton and Laroque (1992) used the basic instruments $[1, p_{t-1}, p_{t-2}, p_{t-3}]$ and a suboptimal weighting matrix which would be optimal if disturbances were not heteroskedastic.[5] Although this format has been standard in the literature, a few concerns are natural even without any econometric investigation. First, since the optimal weighting matrix is justified at least asymptotically, it would be natural to ask if application of it to estimating storage model can be helpful. Second, being essentially a Markovian model, given lag one price, previous information is irrelevant. Then, why include lag one to three in the instrument at the same time? Third, theory predicts, as we have seen above, that the conditional variance function is (weakly) increasing in price. Then, doesn't a decreasing function of lag price provide better weigh of the residual than the lag price itself? Formally, I ask, at small sample size:

1. If iterative use of estimated optimal weighting matrix is helpful?
2. If including in the instruments prices with more than one lag is necessary?
3. What simple transformations of lag prices perform well?

The answers will jointly point out that BIV is a good preliminary estimator.

Before we proceed, it is important to clarify that the relation between the quality of preliminary estimation and that of the final stage is not clear. As we have seen, the true OIV was not performing the best. Nevertheless, searching for a "well-biased" preliminary

[4] Of course, misspecification in ML estimation is a concern. It is difficult to gauge its effect on the estimation of the inverse demand function without serious investigation. Fortunately, because the proposed method works only with the instruments, serious misspecification will not alter consistency.

[5] Chambers and Bailey (1996) used up to lag 6 and 12 and suboptimal weighting matrix.

estimator seems not a promising question. We proceed with the belief that a reliable preliminary estimator is in general beneficial; after all, it is useful by itself. Appendix C contains some extra simulations about this issue.

1.5.1 The Estimated Optimal Weighting Matrix

I begin with the one-step estimator of Deaton and Laroque (1992) (denoted as DL1), and three two-step GMM estimators using estimated optimal weighting matrix with one to three times of iterations (denoted as DL2, DL3 and DL4, respectively). I also include a continuous-updating estimator (denoted as CU), developed in Hansen, Heaton and Yaron (1996), with the same instrument as DL1 and DL2, and a much simpler just-identified estimator with instrument $[1, p_{t-1}]$ and identity weight.

From Table 1.3, iterative use of the estimated optimal weighting matrix undermines estimation of the cutoff price by producing more downward biases without improving γ much. In contrast, the much simpler just-identified estimator with instrument $[1, p_{t-1}]$ and identity weighting matrix (see last column of Table 1.3) provides more reliable estimates of p^* than any other estimator, and comparable performance in estimating γ.

Table 1.3: The performance of estimated optimal weighting matrix

		DL1	DL2	DL3	DL4	CU	$[1, p_{t-1}]$
	Mean	1.0551	1.0532	1.0535	1.0537	1.0536	1.0552
	25%	1.0397	1.0382	1.0385	1.0387	1.0386	1.0401
γ	50%	1.0522	1.0502	1.0504	1.0506	1.0506	1.0523
$(= 1.05)$	75%	1.0675	1.0651	1.0654	1.0655	1.0654	1.0672
	STD	0.0215	0.0209	0.0214	0.0215	0.0212	0.0210
	RMSE	0.0222	0.0211	0.0216	0.0218	0.0215	0.0216
	Mean	112.2180	110.4749	110.0101	109.8623	114.8467	112.1399
	25%	106.1238	104.5170	103.9877	103.8366	104.3398	106.2537
p^*	50%	111.4010	109.6509	109.1442	108.9626	109.9735	111.3853
$(= 114.1243)$	75%	117.4835	115.5890	115.1465	115.0076	116.8952	117.2435
	STD	8.5436	8.4473	8.6054	8.6835	23.9268	8.2960
	RMSE	8.7533	9.2015	9.5380	9.6727	23.9365	8.5297

Note. 10000 replications with sample size 100.

Our observation is consistent with the finding of, for example, Hansen, Heaton and Yaron (1996), which under the context of CCAPM suggests that further iterating the second step of the two-step GMM estimator can undermine estimation.

This problem was attributed to the correlation between the estimated optimal weighting matrix and error terms in the second step at finite sample size (Altonji and Segal, 1996). The estimated optimal weighing matrix is calculated using previous estimates from the same sample used in the second step and thus correlation occurs.[6]

[6] Similar problem exists for 2SLS estimators: the fitted variables from the first stage correlate with

Two solutions were proposed. The first is the split sample estimator, using different segments of the sample to estimate the moments and the weights. It was proposed by Angrist and Krueger 1995 for 2SLS estimators and adapted to GMM by Altonji and Segal (1996). The second is the jackknife estimator, proposed by Angrist, Imbens and Krueger (1999) originally for 2SLS; Donald and Newey (2000) showed that the CU estimator has a jackknife interpretation.

The split sample approach is not applicable here: the sample size is already too small. Neither does the jackknife approach, as evident by poor performance of CU in estimating p^* from Table 1.3. This provides some support to this chapter's emphasis on choosing appropriate instruments to improve small-sample estimation.

This emphasis is also reasonable from asymptotic perspective: the GMM estimator attaining the efficiency bound for given conditional restriction by choosing optimal instruments is efficient relative to a GMM estimator attaining the efficiency bound for given unconditional conditions by choosing the optimal weighting matrix (Davidson and McKinnon, 1993, p.604). When the optimal instrument implies a just-identified estimator as in our case, the weighting matrix is irrelevant as long as it is positive definite and the derivative of sample moments w.r.t. parameters has full rank. Therefore, at least for large sample size, as far as parameter estimation is concerned, weighting matrix is not important.

1.5.2 Lag length in Instruments

I compare performance of instruments that include prices with various lags. In these estimations, a one-step estimator with sub-optimal weighting matrix in the format of Deaton and Laroque (1992) is used except for the just-identified case, where identity weighting matrix is used.[7]

From Table 1.4, we see that including additional lags undermines estimation. The RMSE is increasing in lag length for both parameters. The 25 and 75 percentile values suggest that the distribution of the estimator is more concentrated around the true for fewer lags.

In the context of CCAPM, Tauchen (1986) found strong evidence at sample size of 50 and 75 that the mean bias increases with an increasing lag length of instrument. His finding was largely confirmed by Kocherlakota (1990), Ferson and Foerster (1994), and Hansen, Heaton and Yaron (1996). Here, adding more lags seems to bring more dispersion without affecting mean biases much.

From an asymptotic view when disturbance displays conditional heteroskedasticity or serial correlation, use of additional instruments typically delivers increased asymptotic efficiency. But, first, this does not necessarily apply to small-sample case. For example, West et al. (2009) proposed an instrumental variable estimator of linear models

the error terms in the second stage at finite sample size (Angrist and Krueger, 1995; Angrist, Imbens and Kruger, 1999).

[7] By the result above, iteration with estimated optimal weighting matrix is not useful.

Table 1.4: The performance of instruments with various lag lengths

Lag length:		1	2	3	4	$[1, p_{t-2}]$
	Mean	1.0552	1.0550	1.0551	1.0551	1.0655
	25%	1.0401	1.0398	1.0397	1.0396	1.0352
γ	50%	1.0523	1.0521	1.0522	1.0521	1.0561
(= 1.05)	75%	1.0672	1.0674	1.0675	1.0677	1.0848
	STD	0.0210	0.0214	0.0216	0.0219	0.0492
	RMSE	0.0216	0.0220	0.0222	0.0225	0.0516
	Mean	112.1399	112.2429	112.2180	112.2122	114.8727
	25%	106.2537	106.2017	106.1238	106.1151	100.3094
p^*	50%	111.3853	111.4456	111.4010	111.3925	108.5347
(= 114.1243)	75%	117.2435	117.4756	117.4835	117.4499	121.3292
	STD	8.2960	8.5038	8.5436	8.6040	24.3166
	RMSE	8.5297	8.7090	8.7533	8.8135	24.3269

Note. 10000 replications with sample size 100.

with conditionally heteroskedastic disturbances to "exploit information in all lags of instruments, unconstrained by degrees of freedom limitations". Their simulation indicated large finite-sample efficiency relative to alternative estimators for sample size over 1000. The performance, however, deteriorates dramatically when sample size reduces to 250. The analysis here provides evidence of efficiency loss of using instruments with lag length over one at sample size of 100, for a nonlinear Markovian model with conditionally heteroskedastic, serial uncorrelated disturbances.

Second, adding more lags does not necessarily increase efficiency. There is in fact a small literature, related to the concept of "redundant instruments" of Breusch, et al. (1999), describes how efficiency gains may result from using finite lags (Kim et al., 1999; Broze et al., 2001). A set of instruments is redundant relative to a given set of instruments if adding the former to the latter does not increase asymptotic efficiency. West (2002) provided an integrate treatment of Kim et al. (1999) and Broze et al. (2001), and emphasized that whether there is efficiency gain from using more lags depends on the form of conditional heteroskedasticity. Unfortunately, their result does apply here directly.[8] More importantly, the result instrument optimal in the subclass they considered still requires preliminary estimation, which comes back to the situation we face here.

To obtain some insights on this problem, assume the cutoff price is known, then sample

[8]Particularly, it can be proved that a requirement of theorem 2 of West (2002), $Eu_t^2 u_{t-k} u_{t-j} = 0,\ j \neq k$, is not satisfied in the storage model.

moment becomes:

$$\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}(p_t-\gamma\min\{p_0^*,p_{t-1}\})\cdot 1\\(p_t-\gamma\min\{p_0^*,p_{t-1}\})\cdot p_{t-1}\\\vdots\\(p_t-\gamma\min\{p_0^*,p_{t-1}\})\cdot p_{t-s}\end{bmatrix}, \tag{1.23}$$

Thus,

$$BIAS=\left(G^{'}WG\right)^{-1}G^{'}W\Psi\left(\theta_0\right) \tag{1.24}$$

where $G=T^{-1}\sum_{t=1}^{T}\min\{p_0^*,p_{t-1}\}\otimes[1,p_{t-1},\ldots,p_{t-s}]^{'}$ and W is weight matrix. The convenience of this simplification lies in that its right hand side does not contain the estimator.[9] Then, distribution of the bias is simply that of the second component on the right hand side.

Figure 1.6 plots the distributions of the bias for instruments with various lag length when sub-optimal weighting matrix is used. It is illuminating to see that the four distributions are indistinguishable. This strongly suggests that at small sample size, when lag-one price is already in the instrument, adding older prices is meaningless. This seems echoing the very implication of Markovian models: given latest information, older information is irrelevant (though they are valid instruments).

Also note that at small sample size the numerator in the bias formula above is not small. This implies that if the probability of nearly-zero denominator is relatively big, bias can be severe. Simulation just showed that including less relevant instruments deteriorates the distribution of the small-sample bias by introducing more dispersion. It is then reasonable to conjecture that, at small sample size, less relevant (in the limit) instrument not only implies lower mean of but also more volatile correlation coefficients between it and the variable to be instrumented. Then, if correlation coefficient is a measure of how useful an instrument is, such property of less relevant instrument suggests that the extra instability it introduces can easily offset its limited contribution (if there is).

Specifically, p_{t-s}, $s\geq 2$ is a valid instrument and relevant in the mean or the limit sense.[10] But, at small sample size, when the correlation coefficient specified above is far from converging, its dispersed distribution renders its mean not informative. This is strongly confirmed by Figure 1.7 which presents the distributions of correlation coefficient between p_{t-s} and $\min\{p_0^*,p_{t-1}\}$ for s from 1 to 4. It can be seen easily that less relevant

[9] A formal treatment of the problem requires approximation to the finite sample bias. At large sample size, this can be done by applying the Mean Value Theorem to the first-order condition of the estimation problem, and relying on consistency for simplification. It is difficult to do so at finite sample size in nonlinear cases because the estimator enters the approximation expression. For linear regression, Buse (1992) derived an expression for the approximate bias of instrument estimator at finite sample.

[10] In GMM, if the instruments Z_t are relevant, then $E\left[u_t\left(\theta\right)\otimes Z_t\right]\neq 0$ for $\theta\neq\theta_0$. If it is nearly 0 for $\theta\neq\theta_0$, then θ is weakly identified (Stock, Wright and Yogo, 2002). Even the lag-three price does not suffer from weak identification with plentiful data. It can be easily shown that the large-sample approximant to $E\left[u_t\left(\theta\right)p_{t-3}\right]$ around the true parameter has significant negative gradient in both directions. For details see Appendix B.

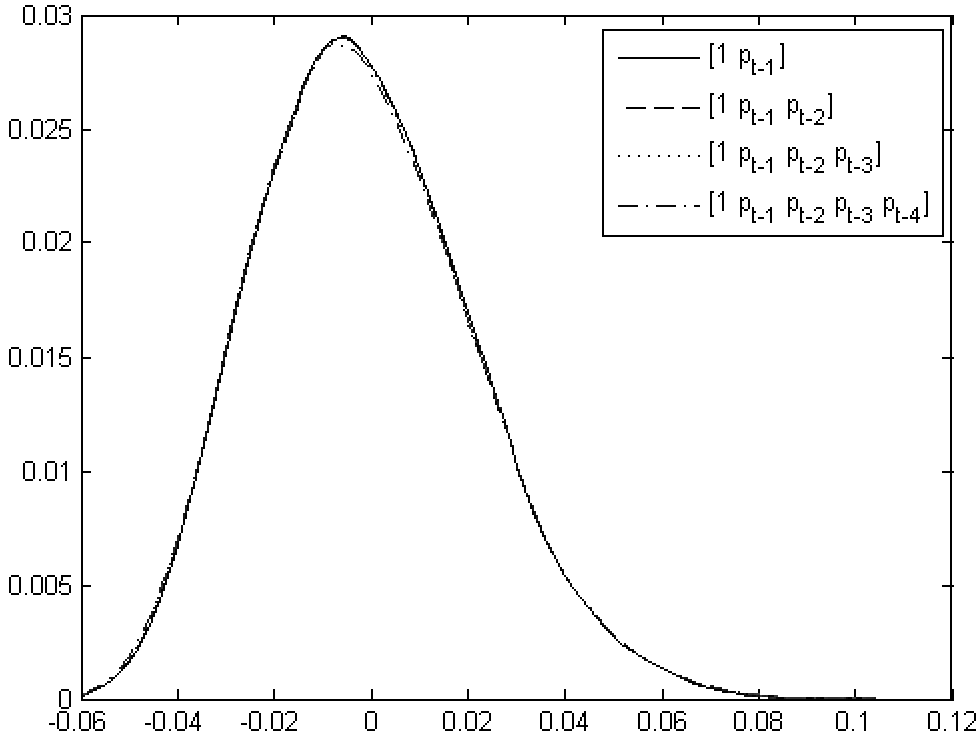

Figure 1.6: Bias of γ if p^* is fixed at its true value vs. lag length in instrument

instrument tends to have lower mean but much larger deviation and therefore higher chance of getting near 0.

Lacking formal treatment, this analysis justifies at small sample size the use of small set of instrument and the necessity of treating instrument relevance as a distribution rather than just a mean value.

1.5.3 Transformation of Instruments

To be focused, I only investigate transformations of instruments of lag one price. Specifically, I compare the performances of instruments that contain various powers of p_{t-1} and a constant one.

From Table 1.5, the performance in estimating γ improves as the power of lag one price decreases from 2 to -1.5: the RMSE for γ is reduced by 21%. Other statistics also improve as the power falls from 2 to -1.5: The mean bias reduced by a dramatic 51.7% and median bias is reduced by an incredible 95.8%. The trend in RMSE, however, is reversed when the power reduces further -2, as evident by the significant increase in value.

Table 1.5: The performance of instruments with various various powers of lag one price

power of p_{t-1}:		2	1	0.5	-0.5	-1	-1.5	-2
	Mean	1.0558	1.0552	1.0548	1.0544	1.0542	1.0528	1.0505
	25%	1.0394	1.0401	1.0403	1.0406	1.0408	1.0396	1.0384
γ	50%	1.0524	1.0523	1.0521	1.0518	1.0518	1.0501	1.0463
$(= 1.05)$	75%	1.0685	1.0672	1.0664	1.0652	1.0649	1.0622	1.0543
	STD	0.0228	0.0210	0.0202	0.0190	0.0185	0.0183	0.0220
	RMSE	0.0235	0.0216	0.0208	0.0195	0.0190	0.0185	0.0220
	Mean	112.067	112.140	112.198	112.285	112.316	113.044	115.128
	25%	105.809	106.264	106.417	106.624	106.687	107.293	107.527
p^*	50%	111.229	111.385	111.543	111.632	111.689	112.362	111.692
$(= 114.124)$	75%	117.417	117.244	117.259	117.264	117.405	118.240	121.253
	STD	8.7839	8.2960	8.1322	7.9577	7.9342	8.3219	11.0701
	RMSE	9.0213	8.5297	8.3578	8.1672	8.1373	8.3219	11.1149

Note. 10000 replications with sample size 100.

As to p^*, from power 2 to -1 the incremental improvement is diminishing: when the power falls from 1 to 0.5, the RMSE falls by more than 0.17, while when it falls from -0.5 to -1, RMSE reduces by merely 0.03. Like the case for γ, there seems to have a turning point: when the power falls to -1.5, the RMSE increases to about the level at power 0.5; when the power further reduces to -2, the RMSE becomes even bigger than at power 1. Other statistics exhibit the same trend: they all improve until the power moves further negative from minus one.

Figure 1.8 repeats the exercise of Figure 1.6 for instruments of various powers of lag one price the distributions of small-sample bias in estimating γ when the cutoff price

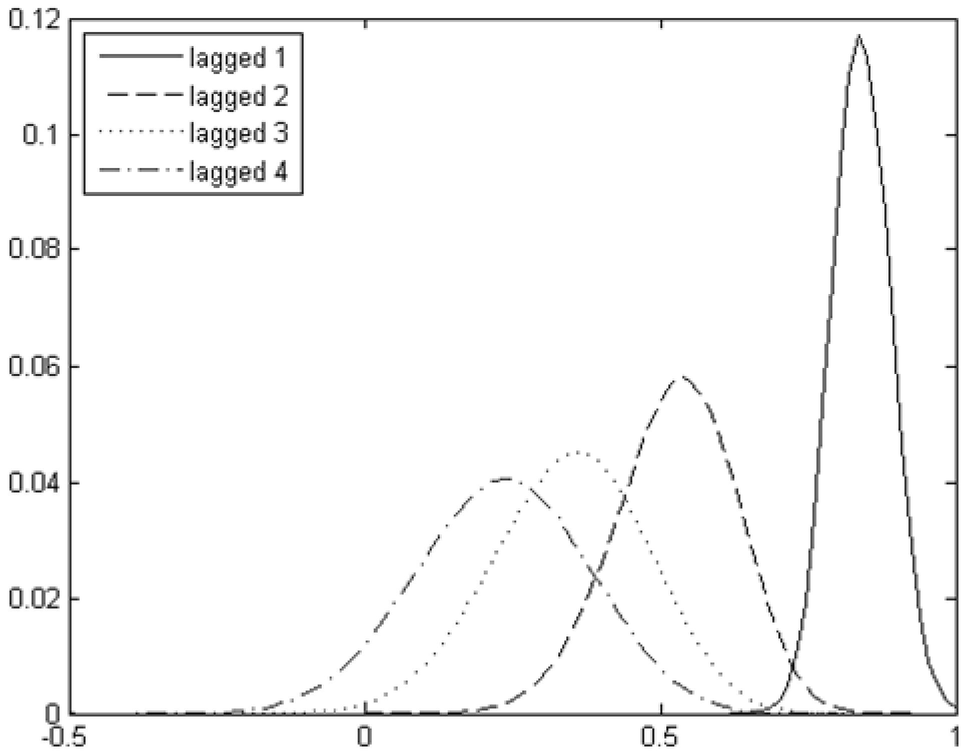

Figure 1.7: Lag vs. correlation coefficient as measure of instrument relevance

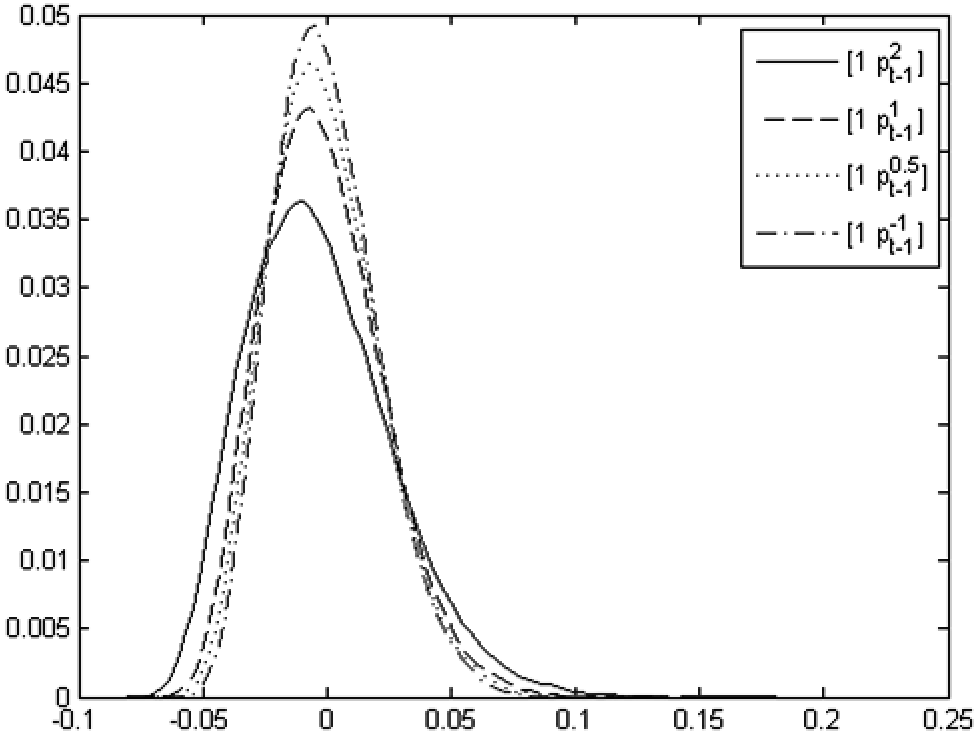

Figure 1.8: Bias of γ if p^* is fixed at its true value vs. powers of lag one price in instrument

is fixed at the true. As the power decreases from 2 to -1, the distribution becomes more concentrated around 0: when the power is 2 the distribution has mean -0.0029 and standard deviation 0.0287, but when it reduces to -1 the mean becomes -0.0006 and the standard deviation becomes 0.0206. This general pattern is consistent with estimation of γ in Table 1.5. Obviously, a very simple transformation of basic instrument can yield dramatic improvement in estimation at small sample size, but at the same time tricky because the effects are not monotonic and the optimal level of transformation is generally unknown.

If there is an optimal power of lag one price in the class of instruments considered, then under the current parameterization it should be between -1.5 and -0.5 for p^* and between -1 and -2 for γ. It is far beyond the scope of this chapter to find the exact optimal power for given parameterization. Even if we know it, it most likely will contain unknown part that requires preliminary estimation and then we come back to the current situation. Therefore, practically, it would be reasonable to just find a specific simulation-proved power that works satisfactorily within a range of parameterizations.

Moreover, the complication is not as serious as it seems. As mentioned above, Table 1.5 suggests that, for example, the estimation of p^* improves at a diminishing rate when the power decreases to -1. Thus, the performances at power -1 and at the optimal power may be close for estimating the cutoff price. Therefore, -1 may work well enough in practice. This justifies the use of $\left[1, p_{t-1}^{-1}\right]$ in the preliminary estimator in section 4, and of course it is also well-performing itself.

The dramatic change in performance due to a simple transformation of instrument requires further discussion. Usually, two conditions, validity and relevance, of instruments receive the most attention. But, instrument estimation also accommodates the idea of weighting the residuals. In the storage model, conditional variance of disturbance is weakly increasing the lag price, then using the lag one price as instrument means residual with large conditional variance is up weighted, while using the inverse lag one price means it is down weighted.

Recall that the basic insight of weighted least squares is to give less weight to the less precise measurements, then the advantage of inverse lag one price as instrument over lag one price itself is not difficult to understand. In fact, simple derivation will reveal that the first order condition of the nonlinear least squares with the reciprocal conditional variance of the disturbance as weight is the sample moment of the GMM estimator with the optimal instrument. Of course, this optimal instrument is infeasible and this chapter just offered a method to approximate it.

1.6 Conclusion

This chapter exploits the potential of GMM estimators of the rational expectations competitive storage model by investigating the small-sample performance of various instruments. I propose a method to approximate the infeasible optimal instruments, which is shown numerically to bring small-sample efficiency relative to the benchmark GMM estimator in the empirical literature of storage model. I also identify a reliable simple estimator that requires no preliminary estimation. This work is practically relevant and interesting from a methodological perspective.

This chapter is simulation-based. A few interesting observations lack formal treatment. The procedure to choose the auxiliary model and the relation to the final performance is generally unknown. The rescaling technique in this chapter is validated by the existence of an estimable kink on the conditional variance function. Application of the technique to other models requires special knowledge to find an estimable reference point as a benchmark for rescaling.

Chapter 2

Price Behavior under a Trending Commodity Storage Model: Theory, Simulation and Empirical Evidence from the World Corn Market

2.1 Introduction

This chapter develops a model to understand the dynamics of commodity prices specifically considering the trend in price. This topic is of great policy and welfare relevance. For example, it is relevant to development and food security especially during a time of expansion of biofuels, as many less-developed countries depend heavily on exports of a few primary commodities while at the same time being vulnerable to surges in the prices of foods. A better understanding of the behavior of commodity price is essential for the formulation of trade policies, agricultural policies, and the designs of mechanisms for risk-sharing and warning systems for potential food price spike.

This model, in contrast to previous literature, is able to fit the empirical patterns in prices and resulting inventory fluctuations quite well and, unlike existing models, is applicable to important questions involving structural changes in commodity markets. It highlights the role of speculation in explaining the short-run dynamics of prices. The estimation results obtained by fitting this model to the actual price data are more reasonable than those in the literature, and useful to financial and policy analysis.

The behavior of speculators can explain many important features of commodity prices. By buying cheap and selling dear, commodity speculators can induce positive autocorrelation in price, a stylized fact in the historical price data. The inventories they carry have an asymmetric effect on price, which explains the skewness and kurtosis in the empirical price distribution. These inventories may run out (i.e., a stockout may occur) after a row of bad harvests, leaving the market unprotected. This explains the infrequent but violent price spikes observed in most commodity markets. Given these observations, it would be surprising if speculation is irrelevant to the short-run dynamics of commodity price.

However, there is not yet a consensus among economists. Deaton and Laroque (1992, 1995, and 1996) and others (e.g., Ng and Ruge-Murcia 2000) argue that the behavior of risk neutral rational expectations speculators is "incapable of generating the high degree of serial correlation of most commodities", unless the autocorrelation coefficients of production are "almost as large as the autocorrelation in prices themselves". They also find that the expected real rates of return on storage implied in the commodity markets "are all implausibly small".[1] One consequence of such disappointment is Deaton and Laroque's (2003) turning "away from inventory as an explanation for the short-run dynamics of prices". This shift in understanding of price behavior might well influence policy responses to relevant issues including those mentioned above.

I re-investigate this issue and find that the previous literature may be missing an important factor.[2] The claim regarding the autocorrelation in price is based on comparison of the implications of the trendless speculative model with the properties of price samples

[1] The response of the literature to this problem was passive. In much of the literature, the real rate is assumed at a heuristic level rather than estimated. For example, it is fixed at 5% in Deaton and Laroque (1995, 1996).

[2] For other problems with the numerical and empirical methods of Deaton and Laroque (1995, 1996) see Cafiero, et al. (2010, 2011).

that most likely contain a downward trend (see Figure 2.1).[3] Their inference regarding the expected real rate of return is based on fitting a trendless econometric model to the likely trending data. Their approach may be internally inconsistent if a trend exists because commodity speculators with rational expectations cannot ignore the price trend when making arbitrage decisions, and so the trend must reveal itself in the price dynamics.

I show that ignoring even a mild downward trend in only a fraction of the sample can significantly upward bias the sample autocorrelation coefficient in price and, downward bias the implied expected real rate of return. On the other hand, speculation recognizing a trend is able to produce price autocorrelation, estimated ignoring trend, that reaches the previously observed level, without any serial dependence in production other than an upward trend. In contrast, if speculation is absent, a reasonable level of price trend induces limited price autocorrelation, estimated also ignoring trend. These observations suggest that recognizing a real trend can re-establish the role of speculation in explaining the short-run dynamics of prices. This work is a step further from Cafiero, et al. (2011) which was the first successful reexamination of the autocorrelation problem.[4] This work also is the first attempt in the literature to address the problem with inferring the expected real rate of return.[5]

To derive the above results is nontrivial. There are several challenges that I am able to address. First, it is difficult to directly solve for a unique non-stationary equilibrium to a trending dynamic forward-looking model. Second, at limited sample size, the effects of ignoring trend are entangled with small-sample bias. How to separate them? All these concerns are well resolved using both analytical and numerical methods.

I show using market data how I obtain empirical support from the world corn market over the period 1961 - 2005. The expected real rate of return estimated recognizing trend (i.e., 1.86%) falls into the range of long run estimates of riskless real rate of the U.S. economy (i.e., about 2%)[6]. In contrast, if a trend is ignored, the estimate is much

---

[3]Despite the recent rapid ascent of world food prices, the agricultural productivity change that outpaced the growth in population and per capita income has led to decades of downward trend in real prices of many major food commodities. The average real price of corn from 2001 to 2010 was about half of that from 1961 to 1970 (see Figure 2.1). Meanwhile, the world production of corn in 2010 was more than four times of that in 1961. In contrast, the world population was a little more than doubled during the same period. Accordingly, the world per capita production of corn in 2010 was almost as twice as that in 1961 (See Figure 2.2). Therefore, Deaton and Laroque (1992, 1995 and 1996) which used price sample over 1900-1987 ignored the downward trend that likely exists in (at least) the last 27 years of their price sample. Of course, this is somewhat "Monday morning quarterback": when looking only at the sample over 1900-1987, the last 27 years does not necessarily exhibit an obvious downward trend.

[4]Another implication of Cafiero et al. (2011) is that any inaccurate numerical component during the investigation of commodity prices can have consequences for our understanding of price behavior. This lesson will be strictly followed in this work.

[5]The response of the literature to this problem was passive. In much of the literature, the real rate is assumed at a heuristic level rather than estimated. For example, it is fixed at 5% in Deaton and Laroque (1995, 1996).

[6]While it is not truly risk-free, the Treasury Inflation Protected Security (TIPS) is the nearest thing to a safe long-term investment. The 10-year TIPS yield was about 2%.

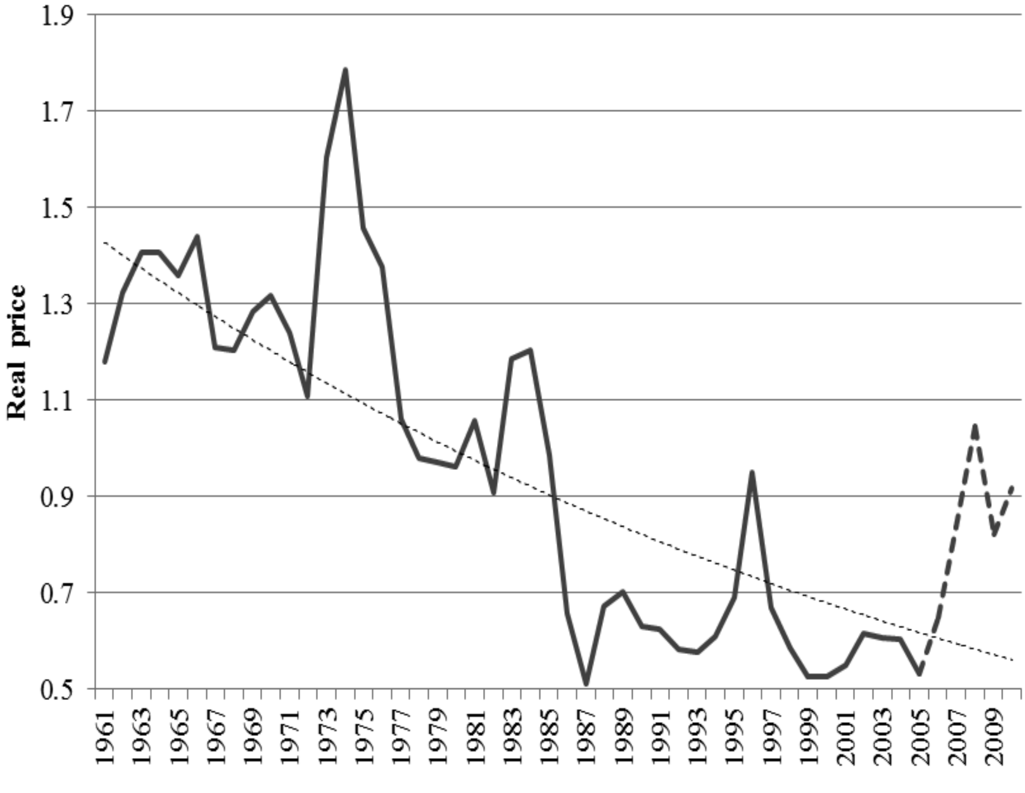

Figure 2.1: Real world prices for corn

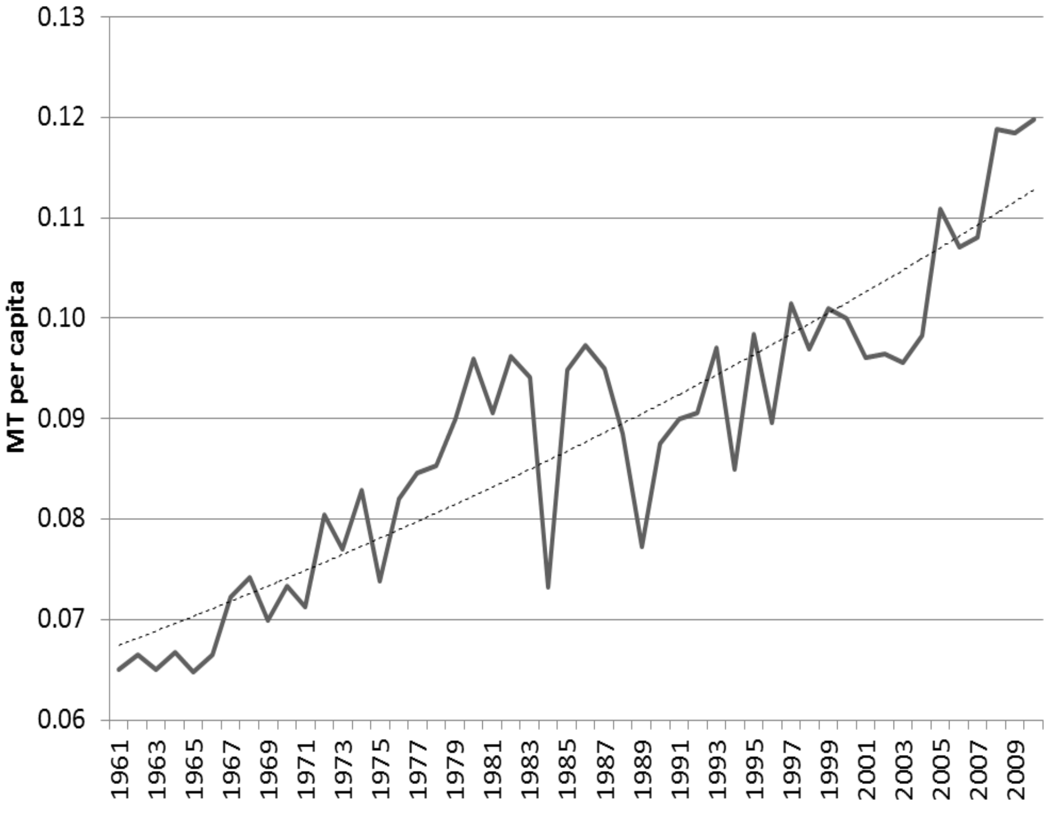

Figure 2.2: Per capita world corn production

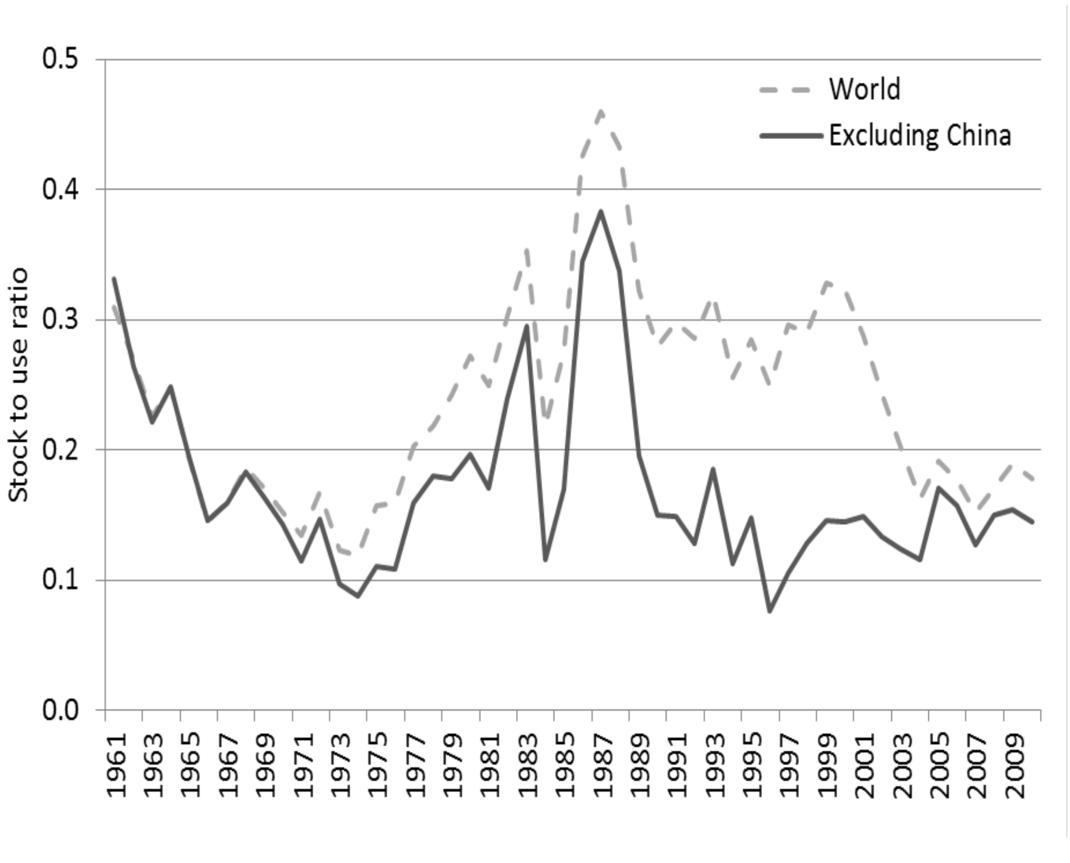

Figure 2.3: Stock to use ratio for world corn

closer to zero (i.e., 0.84%). Ignoring trend is also shown to upward bias the first-order autocorrelation coefficient by 47% and the second-order coefficient by 300%.[7] Finally, ignoring trend suggests no stocksouts in the world corn market beyond the 1970s while recognizing it implies stockouts that are well aligned with historical lows of the stock-to-use ratio (excluding China) in the 70s, 80s and 90s (see Figure2.3).

Section 2 of this paper shows that one way to solve a trending dynamic forward-looking model is to assume a given trend in the underlying stochastic production process, and this trend, for a given demand function, can be factored out of the Euler equation and become trend in price, or vice versa. If the detrended Euler equation defines a unique solution to a stationary model, then the equilibrium of the trending model can be recovered by adding the trend to the equilibrium of the corresponding stationary model. Therefore, the existence of a unique solution to a trending storage model requires the trends in real price and productivity to be mutually determined. This imposes extra restrictions in empirical exercises involving trends if both price and quantity data are to be used.[8]

Under the model developed in section 2, section 3 shows using numerical methods that ignoring the downward price trend creates an artificially high chance that below- (above-) average price induces below- (above-) average price. Simulation shows that this applies when the trend exists in the entire sample or only part of it. Clearly, part of the high observed price autocorrelation can be an illusion from ignoring trend.

Section 3 also addresses the effects of ignoring the price trend on inference of the expected real rate of return. When arbitrage opportunities are fully exploited, forward price expectations are current prices plus marginal storage cost multiplied by a multiplicative factor that greater than unity. Without admitting a convenience yield of storage, the multiplicative factor equals the expected real rate of return adjusted by an (arguably negligible) physical deterioration rate. Section 3 shows that an estimator for the multiplicative factor ignoring the downward price trend may be downward biased.

In section 3, numerical experiments are used to separate the effects of ignoring the trend from those of small sample bias. These experiments confirm that the former is indeed a main source of the previous inconsistency between the theory and observation.

Section 4 and 5 fit the implied autoregression of price to the world corn price from 1961, when the green revolution was about to begin, to 2005, right before the dramatic biofuel expansion, using three Generalized Method of Moments (GMM) estimators that approach the price trend differently. The first estimator works with the autoregression of detrended price and the price trend simultaneously. The second estimator adopts the conventional

[7] If we focus instead on the period over 1900 - 1987, then ignoring the price trend over the sub-sample 1961 - 1987 assuming that earlier data is trendless upward biases the 1st and 2nd order sample autocorrelation coefficients by about 23% and 150%, respectively.

[8] There is a rich literature on the examination of the Prebisch (1950) and Singer (1950) hypotheses. For example, using ample data, Harvey et al. (2010) show that 11 commodities have significant downward trends over all or some fraction of the sample period. The yield trend has been intensively studied as well. This paper combines these two trends with internal consistency under the framework of rational expectations competitive storage.

approach of estimating the autoregressive relation using previously detrended data. The third estimator ignores the price trend and is presented for comparison. The first and second estimators produce similar implications for stockouts but the second yields an implausibly high estimate for the multiplicative factor. A numerical experiment provides support to the finite-sample superiority of the first estimator over the second and helps rule out the estimates of the latter. Finally, the empirical result of the third estimator makes little economic sense.

Unfortunately, while the simultaneous estimator attains smaller standard errors than previous methods using more efficient instruments, the standard error is still too large for conclusive inference at the sample size of merely 45.[9] This motivates the use of the more efficient approach like Maximum Likelihood (ML) and/or incorporating information in the quantity data in the future work.

Section 6 concludes with a summary of results of the paper and a list of limitations.

2.2 A Model of Commodity Storage with Trend

Consider a competitive market for a storable commodity. Time is discrete. There are two types of agents: consumers and storers. Supply response to price is ignored for simplicity. Departing from the classic model of commodity storage, the current model explicitly incorporates productivity change and population growth.

For given consumption demand, the price trend reflects the productivity trend and vice versa. Since price data is used for estimation, we specify the price trend and leave the productivity trend to be determined.

Price trend

We assume that the series of real price p_t contains an exponential trend:

$$p_t = \lambda^t \zeta_t, \quad t \geq 0, \tag{2.1}$$

where ζ_t is stationary and $\lambda \in (0, 1)$ is a constant decay factor. Naive as it seems, this type of trend actually captures the essence of many more sophisticated trends. With a constant decay rate over time, this price trend can be easily estimated.

[9] The instrument for the simultaneous estimator is inspired by Chapter 1 of this thesis which shows that the small-sample performance of GMM estimators of the storage model is sensitive to the choice of the instruments. It shows that the lag prices, the instruments used in previous methods, are inefficient in that they put more weight on price disturbances observed with higher volatility. It proposes an instrument which involves calibrating the unknown optimal instrument at which the asymptotic variance covariance matrix attains the semi-parametric efficiency bound.

Quantity variables

Risk neutral, profit-maximizing storers with rational expectations carry total stocks of X_t from period t to the next with a physical deterioration rate $\delta \in (0,1)$. Storers have access to a perfect capital market at a real interest rate r.

The availability Z_t is the production Y_t plus the carry-in from the last period $(1-\delta) X_{t-1}$, and will be divided into consumption C_t and carry-out X_t, i.e.,

$$Z_t = Y_t + (1-\delta) X_{t-1} = C_t + X_t. \tag{2.2}$$

Equilibrium price functions

Let $c_t \equiv C_t/N_t$ be the per capita consumption at period t, where N_t is the population at period t. Let $f(c_t)$ be the per capita inverse consumption demand. Then, the equilibrium of the decentralized economy can be characterized using a social planer's maximization problem of the sum of expected flows of social surplus:

$$\max_{X_t} \int_0^{Z_t - X_t} f(c_t)\, dC_t + \sum_{j=t+1}^{\infty} \frac{1}{(1+r)^{j-t}} E_{j-1}\left[\int_0^{Zj - X_j} f(c_j)\, dC_j\right] \tag{2.3}$$
$$\text{s.t. } Z_{t+1} = Y_{t+1} + (1-\delta) X_t = C_{t+1} + X_{t+1}, \quad X_t \geq 0, \quad \forall t \geq 0.$$

The first-order conditions are:

$$X_t > 0 \text{ if } 0 = -f(c_t) + \frac{1-\delta}{1+r} E_t f(c_{t+1}), \tag{2.4}$$
$$X_t = 0 \text{ if } 0 \geq -f(c_t) + \frac{1-\delta}{1+r} E_t f(c_{t+1}). \tag{2.5}$$

When $X_t = 0$, everything is consumed (i.e., $C_t = Z_t$).

Let $z_t \equiv Z_t/N_t$ be the per capita availability, then the first-order conditions above can be expressed compactly as:

$$f(c_t) = \max\left\{f(z_t), \frac{1-\delta}{1+r} E_t f(c_{t+1})\right\}. \tag{2.6}$$

Imagine that the equilibrium at period $t+1$ is a function $p_{t+1}(\cdot)$ of the state variable z_t. Let $y_t \equiv Y_t/N_t$ be the per capita production. Then the equilibrium at period t is a function $p_t(\cdot)$ satisfying:

$$p_t(z_t) = \max\left\{f(z_t), \frac{1-\delta}{1+r} E_t p_{t+1}\left(y_{t+1} + \frac{1-\delta}{n_{t+1}}\left(z_t - f^{-1}(p_t(z_t))\right)\right)\right\}, \tag{2.7}$$

where $n_{t+1} = N_{t+1}/N_t$ is the population growth rate from period t to $t+1$.

Let $\gamma \equiv (1+r)/(1-\delta)$. The Euler equation (2.7) implies the following autoregressive relation of prices:

$$\gamma p_t = E_t p_{t+1} \text{ if } p_t \leq p_t^*; \tag{2.8}$$

$$\gamma p_t^* = E_t p_{t+1} \text{ if } p_t > p_t^*, \tag{2.9}$$

where $\gamma p_t^* = E_t p_{t+1}(y_{t+1})$. Expressions (2.8) and (2.9) can be written compactly as:

$$E_t p_{t+1} = \gamma \min\{p_t^*, p_t\}. \tag{2.10}$$

Recall that $p_t = \lambda^t \zeta_t$, thus (2.10) becomes:

$$E_t \zeta_{t+1} = \gamma \min\left\{\frac{p_t^*}{\lambda^t}, \zeta_t\right\}. \tag{2.11}$$

Implications for consumption demand and productivity trend

Following Bobenrieth and Bobenrieth (2010), I assume the inverse per capita consumption demand implied by a Hyperbolic Absolute Risk Aversion utility function:

$$p = f(c) = (A - B \cdot c)^{1/\eta}. \tag{2.12}$$

This function accommodates linear (when $\eta = 1$) and constant elasticity (when $A = 0$) demand.

Given (2.12), the price trend (2.1) implies the following trend in per capita production:

$$y_t = \lambda^{t\eta}\varepsilon_{y,t} + \left(1 - \lambda^{t\eta}\right)\frac{A}{B}, \quad t \geq 0. \tag{2.13}$$

To see this, define:

$$\varepsilon_{z,t} \equiv \lambda^{-t\eta}\left(z_t - \left(1 - \lambda^{t\eta}\right)\frac{A}{B}\right); \tag{2.14}$$

$$\zeta_t(\varepsilon_{z,t}) \equiv \lambda^{-t} p_t\left(\lambda^{t\eta}\varepsilon_{z,t} + \left(1 - \lambda^{t\eta}\right)\frac{A}{B}\right). \tag{2.15}$$

Then, the functional equation (2.7) can be rewritten as:

$$\zeta_t(\varepsilon_{z,t}) = \max\left\{f(\varepsilon_{z,t}), \frac{\lambda}{\gamma}E_t\zeta_{t+1}\left(\varepsilon_{y,t+1} + \frac{1-\delta}{n_{t+1}\lambda^\eta}\left(\varepsilon_{z,t} - f^{-1}\left(\zeta_t\left(\varepsilon_{z,t}\right)\right)\right)\right)\right\}. \tag{2.16}$$

When $\lambda/\gamma < 1$ and $(1-\delta)/n_{t+1}\lambda^\eta$ is stationary, the functional equation (2.16) has a unique solution $\zeta(z)$ by the fixed point theorem. It is easy to see that $\zeta(z_t)$ is the unique *stationary rational expectations equilibrium* of a stationary dynamic model. We call this trendless model the hidden model.

The solution algorithm above involves factoring the productivity trend out of the Euler equation. The process is reversible because of the uniqueness of the solution to the hidden model. A similar technique can be seen in Deaton (1991) in the context of saving under a liquidity constraint. In the literature on storage, productivity change induced price trend was introduced in Bobenrieth, et al. (2010) and homogeneous income growth (more relevant for oil than for grains) by Dvir and Rogoff (2010). The current model emphasizes productivity growth, incorporates population growth and does not consider income growth in modeling.

Testable autoregressive relation of price

It is important to realize that the hidden model and the observable trending model enter stock-outs always at the same time, i.e.,

$$\forall t \geq 0, \quad Z_t - N_t f^{-1}\left(p_t\left(z_t\right)\right) = 0 \Longleftrightarrow \varepsilon_{z,t} - f^{-1}\left(\zeta_t\left(\varepsilon_{z,t}\right)\right) = 0. \tag{2.17}$$

That is,

$$p_t^* = \lambda^t \zeta_t^*, \quad t \geq 0,$$

where ζ_t^* is the cutoff price of the hidden model. This is useful because (2.11) can now be rewritten as:

$$E_t \zeta_{t+1} = \gamma \min\left\{\zeta_t^*, \zeta_t\right\}. \tag{2.18}$$

If $\varepsilon_{y,t+1}$and the deviation from the constant expected growth rate in population are further assumed to be *i.i.d.*, ζ_t^* will be constant over time, i.e., $\zeta_t^* \equiv \zeta^*$. Then, when price can be detrended properly, we are able to estimate γ/λ and ζ^* using the limited information estimators such as the GMM. Since λ is known after detrending, we can obtain estimate for γ and recover the trending cutoff price of the observable model by adding trend to ζ^*.

Throughout the rest of the paper, the *i.i.d.* assumption above will be maintained. There are two reasons for this assumption. First, since we want to understand how the introduction of a price trend in the storage model can affect the autocorrelation in price, it is reasonable to assume away the factors in the production process other than an upward trend that can also contribute serial dependence in price. Second, if serial dependence in $\varepsilon_{y,t+1}$ is introduced, multiple cutoff prices will be implied (see Deaton and Laroque 1995, 1996); inferring all these cutoff prices using limited information estimators requires plentiful of data, which is unrealistic in the current situation, and ad hoc assumptions on the autocorrelation in $\varepsilon_{y,t+1}$.[10]

Finally, noting that (2.18) does not depend on the specifications of the demand and the productivity trend, estimation with (2.18) is robust to those specifications. If, however,

[10]By assuming a periodic harvest process, Chambers and Bailey (1996) estimated a storage model with multiple cutoff prices using a GMM estimator. Deaton and Laroque (1995, 1996) estimated a storage model with a first-order Markovian harvest using a Pseudo ML estimator. Both the state variable and the transition matrix of the harvest process were specified before the actual estimation.

quantity data were ever to be used, specifications of demand and productivity trend become crucial. For example, when $A = 0$, exponential price trend reflects exponential per capita production. This trend is restrictive in that it requires the level and the volatility of per capita production to grow at a common rate. From Figure 2.2, while there seems a mild exponential trend in the level, the volatility seems to stabilize over the recent decades. Despite its common use in the literature, the exponential trend in the per capita production may not be a satisfactory approximation.

2.3 Empirical Implications of Ignoring the Price Trend

2.3.1 Estimating Autocorrelation in Price Ignoring Trend

Define the k-th sample autocorrelation coefficient for sample $\{p_t\}_{t=0}^{T}$ as:

$$\frac{\sum_{t=1}^{T-k}(p_t - \bar{p})(p_{t+k} - \bar{p})}{\sum_{t=1}^{T}(p_t - \bar{p})^2}, \quad k = 0, 1, 2, \ldots, K. \tag{2.19}$$

Writing in this way, I assume that the applied econometrician ignores the price trend. To see this, note that the population autocorrelation coefficient for the price process p_i between time s and t is defined as:

$$\frac{E\left[(p_t - \mu_t)(p_s - \mu_s)\right]}{\sigma_t \sigma_s}. \tag{2.20}$$

where μ_i and σ_i is the mean and standard deviation of the price at time i, respectively. When p_t is exponentially downward trending, i.e., $p_t = \lambda_t \zeta_t$ where ζ_t is stationary, we have $\mu_t = \lambda^t E\zeta_t$ and $\sigma_t = Std(\zeta_t)$ (Std stands for standard deviation), and (2.20) becomes:

$$\frac{E\left[(\zeta_t - E\zeta_t)(\zeta_s - E\zeta_s)\right]}{Std(\zeta_t)\, Std(\zeta_s)}. \tag{2.21}$$

which is the population autocorrelation coefficient of the detrended price ζ_t. That is, the population correlation efficient of the trending price and its detrended counterpart should be the same, and thus to estimate the population correlation coefficient, proper detrending must be done.

If, however, the applied econometrician fails to recognize the trend in price, he will use (2.19) to estimate the population autocorrelation coefficient. This estimate, of course, will not converge to what he would think. We will see that applying (2.19) to the downward trending price implied by the storage model in the last section can create illusion of high autocorrelation coefficient in price even when the downward trend is weak.

From the last section, the trending price can be generated by first numerically solving the hidden model and then adding the trend to the sequence of stationary prices implied by the hidden model. We have seen that, for a given detrended harvest process, such generated price is equivalent to the trending price generated by directly feeding corresponding expanding harvest to the trending equilibrium price functions

Specifically, for given initial de-trended availability $\varepsilon_{z,0}$ and numerically represented stationary equilibrium price function of the hidden storage model, we can calculate the implied storage at period 0 for period 1 and the implied equilibrium price ζ_0. Given a random generated de-trended harvest, we can calculate using the dynamic motion of the de-trended quantities to obtain the total de-trended supply of period 1. Then, we calculate the implied storage for period 3 and the implied price at period 2. Iteration with such procedure can generate a sequence of stationary prices. Adding the exponential trend to the stationary price produces the trending prices we are going to work with. Note that the effect of the trend was incorporated into the solution of the associated de-trended model as its discount rate and depreciation rate both contain the decay factor of the trend (see the Euler equation below). For the rest of the section, I will focus on the following parameterizations: the per capita inverse demand function is: $f(c) = c^{-4}$, or the constant price elasticity is -0.25. The ex-ante real rate of return is 2%. For simplicity, I ignore the depreciation rate and storage fee. Let the de-trended harvest shock be *i.i.d.* log-normal associated with the normal distribution with mean 0 and standard deviation 0.1. Assume that the population growth rate is 1%, i.e.,$n = 1.01$. I will experiment with three values of λ: 0.97, 0.98, and 0.99, or 3%, 2% and 1% annual decay rate of real prices.

For given decay rate of price, λ, we first use the standard approach of backward solution to solve the following Euler equation of the hidden trend-less storage model:

$$\zeta_t(\varepsilon_{z,t}) = \max\left\{ f(\varepsilon_{z,t}), \frac{\lambda}{\gamma} E_t \zeta_{t+1}\left(\varepsilon_{y,t+1} + \frac{\lambda^{0.25}}{1.01}\left(\varepsilon_{z,t} - f^{-1}\left(\zeta_t(\varepsilon_{z,t})\right)\right)\right)\right\}. \tag{2.22}$$

Using the procedure mentioned above and for initial de-trended availability that equals the mean harvest, I generate sequences of prices with the designated trend.

Figure 2.4 and Figure 2.5 plots the paths of mean sample auto-correlation coefficients of 5000 simulations for each value of λ and for both trending and corresponding de-trended prices. In Figure 2.4, we can easily see that the paths of the trending prices increase to high levels very fast, while the counterparts of the de-trended prices approach much smaller numbers.

We know that the observed first-order sample autocorrelation coefficients, calculated assuming stationarity, of the major commodities like corn and wheat are in the range between 0.6 and 0.9 and the second-order are in the range between 0.4 and 0.8 (see Table 1 of Deaton and Laroque 1992).

We can easily see in Figure 2.4 and Figure 2.5 that the paths of both the median and mean sample autocorrelation coefficients in the trending prices reach the above observed levels very fast, while those of the detrended prices approach smaller numbers and those of the prices without storage remain at low levels. This observation provides support to the argument that recognizing trend can help externalize the observed autocorrelation.

The illusion of high autocorrelation comes from that the downward trend pushing down the later segment of price below its time average, creating an artifact that below mean price tends to induce below mean price. Since the earlier prices are generally larger in magnitude, while the time average of an exponentially downward trending price becomes small as the sample size increases, it will still remain above some later segments of price in the sample. If the time average is incorrectly treated as a legitimate estimate for the mean of price process, these below-time-average segments of price produce more positive elements in the summation in the numerator of (2.19) than if the downward trend is recognized. Moreover, this upward bias on the sample correlation coefficient from ignored trend is exaggerated by the fact that the denominator of (2.19) is smaller than if the downward trend is recognized.

From Figure 2.4 and Figure 2.5 we also observe, as expected, that stronger trend introduces higher sample autocorrelation coefficients in the trending prices. The order reversed for the detrended prices. This is because stronger trend implies higher deterioration rate of the hidden model, $1 - \lambda^{0.25}/1.01$, and higher discount rate λ/γ, both making the intertemporal dependency of the detrended model weaker.

The previous analysis of the price autocorrelation has focused on the sample period over 1900 - 1987, a large fraction of which might contain negligible trend while the rest exhibits a stronger trend. To take into account of this possibility, I modify the simulation by introducing a trend only starting from the 61st observation.[11] γ is fixed at 1.02 for the entire sample. For simplicity, I assume that the trend is introduced to the market as a surprise. Then until it actually starts the market behaves as if there will not be any structure change.

Counterparts to Figure 2.4 and Figure 2.5 are Figure 2.6 and Figure 2.7. It is surprising that trending price in only a small fraction of the sample can still upward bias the autocorrelation coefficients to a level that is seriously misleading.

From Figure 2.6 and Figure 2.7, the median (mean) sample first-order autocorrelation coefficient given that the partial trend is ignored is upward biased from 0.464 (0.459) to 0.577 (0.558), or about 24%, at sample size of 88 and for the decay factor 0.97. Instead, when the decay factor is 0.98, the distortion for the first-order is from 0.509 (0.494) to 0.578 (0.557), smaller but still about 13-14% upward bias. When the trend is 0.99, the distortion becomes less important (about 0.024 in magnitude).

The result for the second-order autocorrelation coefficient is more striking: about 46% upward bias for the decay factor 0.97, 26% for 0.98, and about 10% for 0.99, for both mean and median at sample size of 88.

The above observations seriously question the legitimacy of the previous challenges against the storage model in terms of implied price autocorrelation.

[11]The U.S. corn yield begins to have an obvious upward trend since about 1940. The trend becomes stronger since around 1960. Strictly speaking, a reasonable simulation analysis requires taking into account of the trend in the 1940s and 1950s. The limited data for this two decades, however, renders inference of the trend over this period unreliable. I assume, for simplicity, that this fraction of the sample has no trend. After all, taking into account of this weaker trend tends to enhance the current argument.

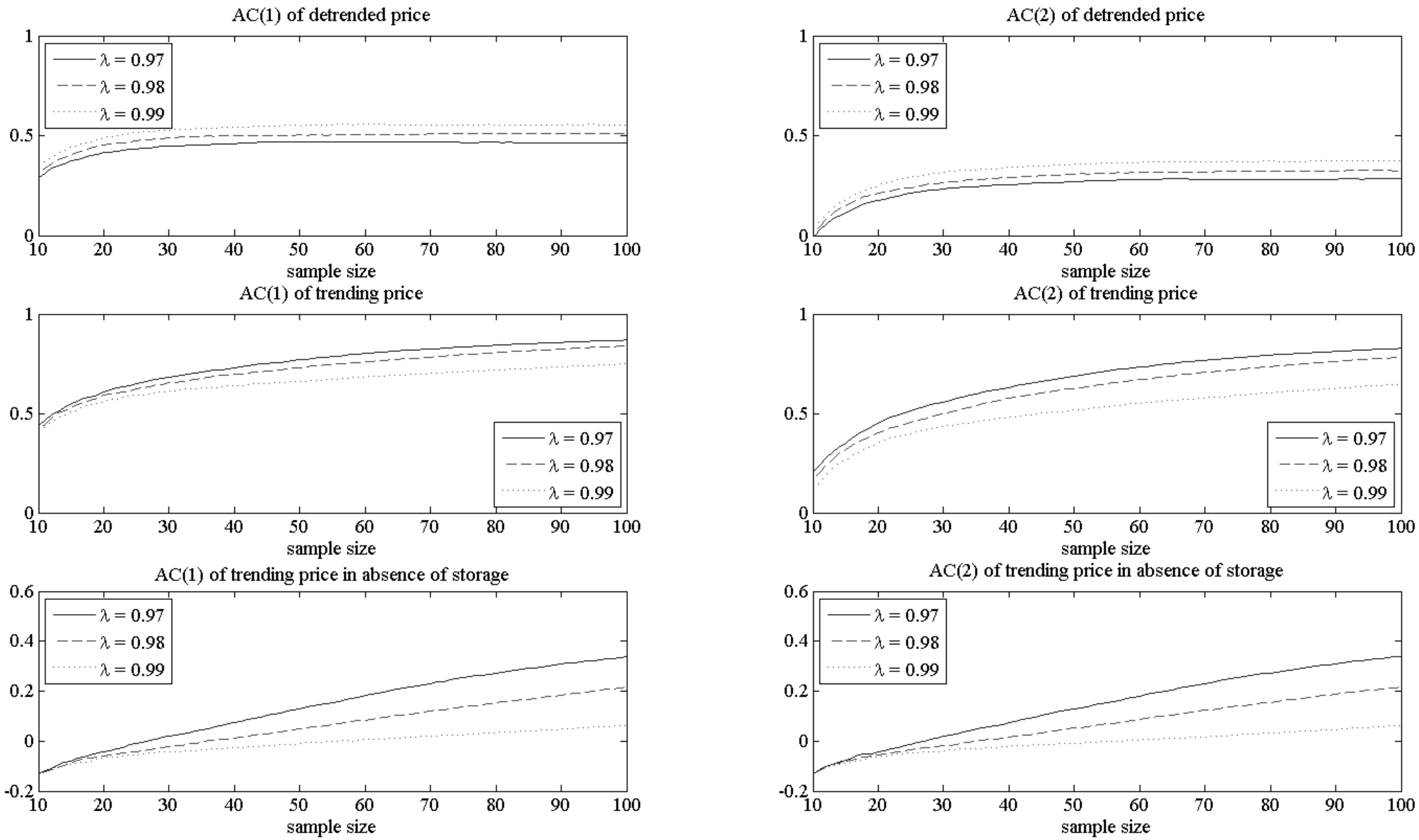

Figure 2.4: The path of median autocorrelation coefficients in simulated price

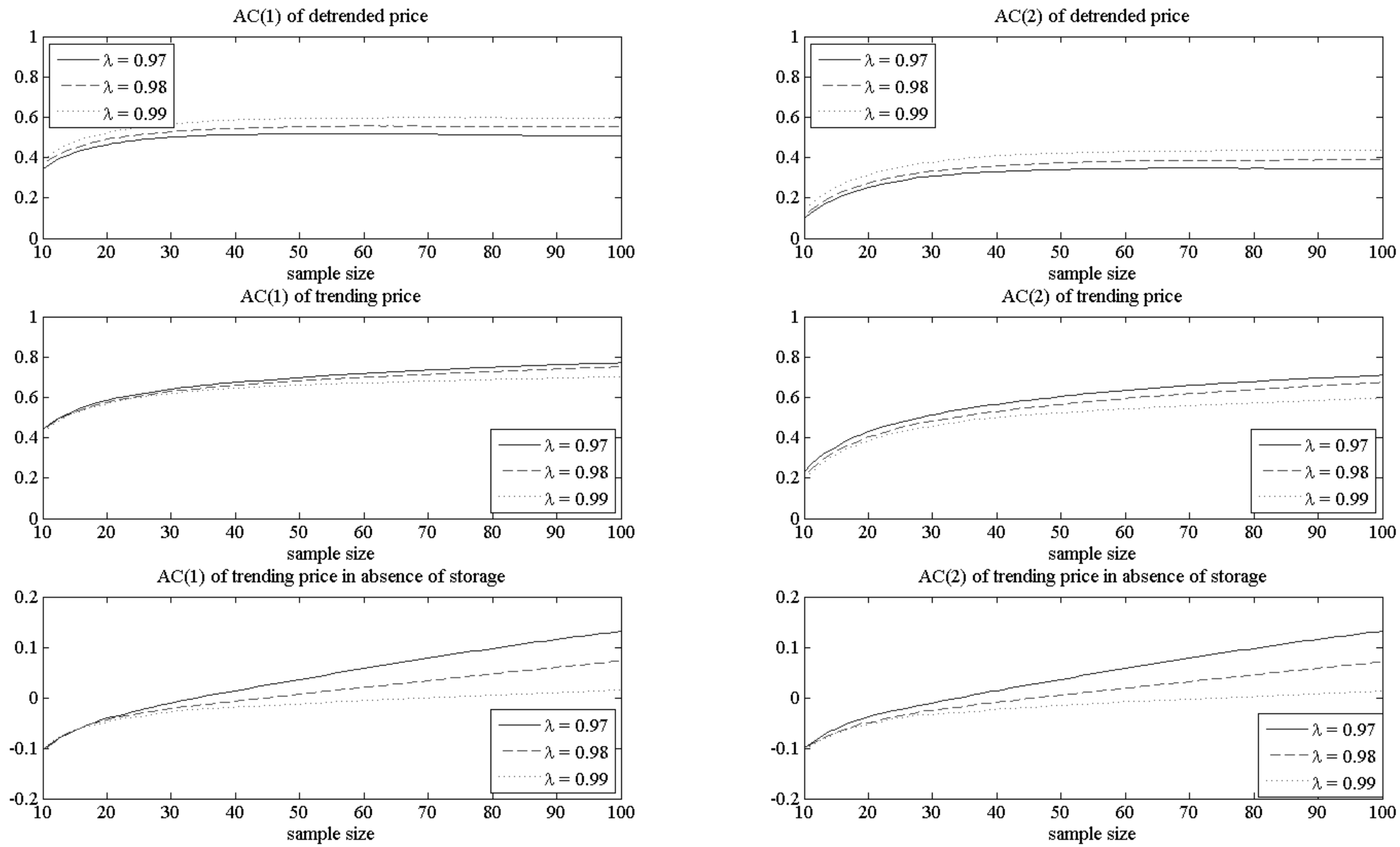

Figure 2.5: The path of mean autocorrelation coefficients in simulated price

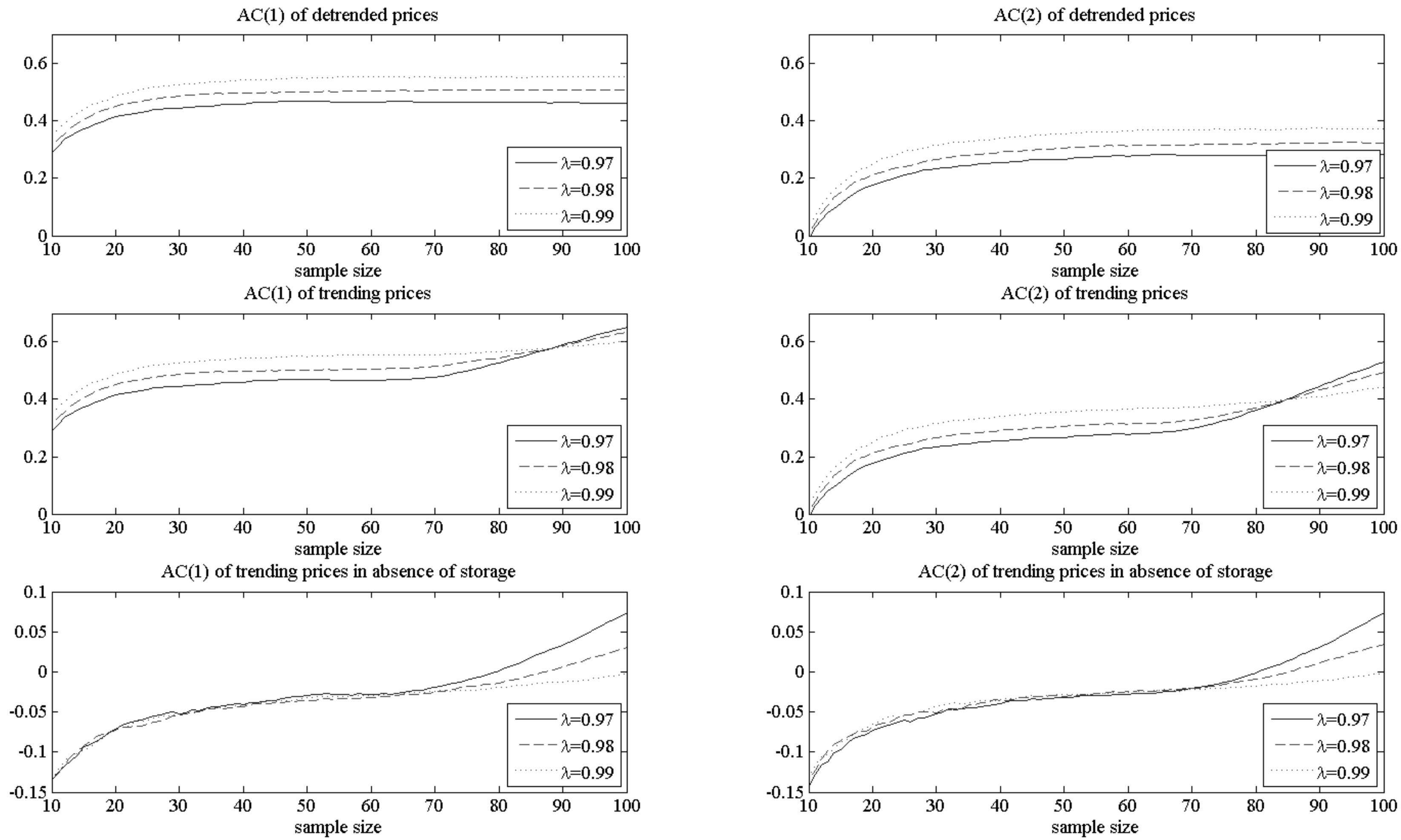

Figure 2.6: The path of median autocorrelation coefficients in simulated price: partial trend

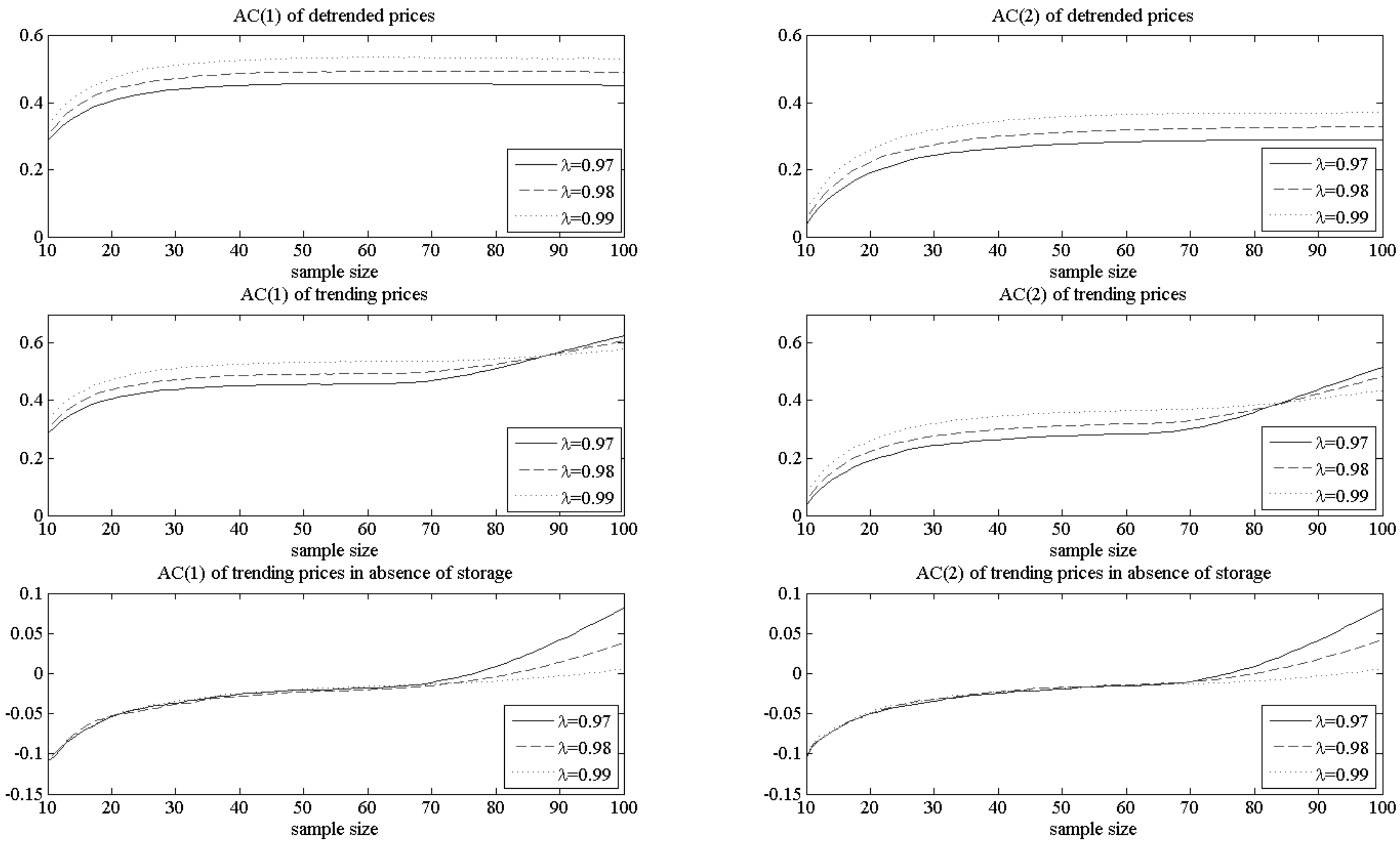

Figure 2.7: The path of mean autocorrelation coefficients in simulated price: partial trend

Understandably, the hypothesis of stationarity would be difficult to reject statistically at small sample size and when the trend is only mild and/or exists in only a small fraction of the sample. Deaton and Laroque (1992) defended their stationarity assumption in the following way:

> "Although we would not wish to claim that all of these prices are stationary, none exhibit any obvious trends, even over such a long period of time."[12]

Should they have had a slightly longer data they would have noticed that the real prices for major grains exhibited an obvious downward trend since at least 1961. The above analysis suggests that ignoring even a mild trend in only a small fraction of the sample can create an illusion of high sample autocorrelation in limited periods. While the basic mechanism creating such an illusion is simple, this small-sample observation is important as it warns of effects that ignoring just a mild partial trend can have on the development of a body of empirical literature.

2.3.2 Estimating the Autoregression of Price Ignoring Trend

If the applied econometrician fails to recognize the downward price trend, he would attempt to estimate γ by minimizing a quadratic form of the following expression:

$$\tilde{u}_{t+1} = p_{t+1} - \gamma \min\{p^*, p_t\}, \quad t \geq 0. \tag{2.23}$$

Working with (2.23) will yield an inconsistent estimator of γ.

To see this, imagine that the applied econometrician attempts to estimate γ using a simple approach of taking average of the ratios between p_{t+1} and $\min\{p^*, p_t\}$. That is, for given $p^*$and given sample $\{p_t\}_{t=0}^{T}$, γ is estimated using:

$$\hat{\gamma}_T = \frac{1}{T}\sum_{t=0}^{T-1} \frac{p_{t+1}}{\min\{p^*, p_t\}}. \tag{2.24}$$

Recall that $p_t = \lambda_0^t \zeta_t$, where λ_0 is the true decay factor. Then (2.24) can be rewritten as:

$$\hat{\gamma}_T = \frac{1}{T}\sum_{t=0}^{T-1} \frac{\lambda_0 \zeta_{t+1}}{\min\left\{\frac{p^*}{\lambda_0^t}, \zeta_t\right\}}. \tag{2.25}$$

For simplicity, we assume that ζ_t is bounded.[13] Then, for given p^*, there exists a finite time N such that the $\min\left\{\frac{p^*}{\lambda_0^t}, \zeta_t\right\}$ equals ζ_t for all $t > N$.

[12] Cafiero et al. (2011) follow Deaton and Laroque (1992, 1995 and 1996) in ignoring any possible trends.

[13] It can be proved that the ergodic set of the total availability of the hidden model has an upper bound, implying a lower bound on the detrended price. For linear demand, for example, the detrended price should also have an upper bound.

Then, we have that,

$$\hat{\gamma} = \frac{1}{T}\sum_{t=0}^{N}\frac{\lambda_0\zeta_{t+1}}{\min\left\{\frac{p^*}{\lambda_0^t},\zeta_t\right\}} + \frac{T-N}{T}\frac{1}{T-N}\sum_{t=N+1}^{T}\frac{\lambda_0\zeta_{t+1}}{\zeta_t}. \tag{2.26}$$

It is easy to see that the first component of (2.26) converges to zero as $T \to \infty$ because the summation in it is bounded from above by the assumed bound over ζ_t.

By the Ergodic Theorem, and the stationary and boundedness of ζ_{t+1}/ζ_t, the second component of (2.26) is converging as well:

$$\frac{T-N}{T}\frac{1}{T-N}\sum_{t=N+1}^{T}\frac{\lambda_0\zeta_{t+1}}{\zeta_t} \xrightarrow{p} \lambda_0 E_\infty\left[\frac{\zeta_{t+1}}{\zeta_t}\right], \tag{2.27}$$

where E_∞ denotes the expectation according to the limit distribution of ζ_{t+1}/ζ_t and $\xrightarrow{p}$ denotes convergence in probability.

Recall that by (2.10) we have:

$$\lambda_0 E_\infty\left[\frac{\zeta_{t+1}}{\zeta_t}\right] = \gamma_0 E_\infty\left[\min\left\{\frac{\zeta_0^*}{\zeta_t},1\right\}\right], \tag{2.28}$$

where ζ_0^* denotes the true cutoff price of the hidden model.

Because $\min\left\{\frac{\zeta_0^*}{\zeta_t},1\right\} \leq 1$ and with positive chance it is strictly smaller than one, we have the probability limit of $\hat{\gamma}_T$ is smaller than the true, i.e.,

$$p\lim_{T\to\infty}\hat{\gamma}_T < \gamma_0.$$

That is, $\hat{\gamma}_T$ is downward biased.

The insight behind the above discussion should extend to more general estimators of (21) and should be valid even when ζ_t is unbounded under certain restrictions on the right tail of the distribution of detrended price.

Noting that Deaton and Laroque (1992) found their estimates of the multiplicative factor to be implausibly small, I conduct a Monte Carlo analysis to study the effects of ignoring a partial trend on the finite-sample performance of their GMM estimator.[14]

Specifically, I compare the distributions of the estimates for γ in the case when there is no trend at all in the entire sample with the cases when the trend starts at the 51st and 61st observation (I assume again that the trend is introduced as a surprise). Such experiment design can help us separate the effects of the small-sample bias from those of ignoring the partial trend. At sample size of 88, the bias of the estimates can be severe even when the econometric model is correctly specified. If the correctly specified estimator still tend to underestimate γ, then the claim that the implausibly small estimates of the

[14]The GMM estimator in Deaton and Laroque (1992) works with the autoregression of stationary price and uses lagged one to three prices plus a constant as instruments.

Table 2.1: Effects of ignoring a partial trend on estimating γ

True	25%	50%	75%
No trend in the entire sample			
1.02	1.0174	1.0219	1.0269
Trend starts at the 51st observation			
1.02	1.0049	1.0089	1.0132
Trend starts at the 61st observation			
1.02	1.0059	1.0100	1.0144

Note. The 25%, 50% and 75% stand for the 25, 50 and 75 percentile values.

multiplicative factor in Deaton and Laroque (1992) are due to ignoring the partial trend in their data is significantly weakened.

In this experiment, the true γ for the entire sample is fixed at 1.02. The harvest is *i.i.d.* standard normal discretized into 10 nodes using Gauss quadrature. The inverse demand function is $p = 10 - b$. The population growth rate is 1.01 for the entire sample. When the trend starts, the decay factor is 0.975. To be consistent with Deaton and Laroque (1992), the sample size is always 88. I simulate 1000 sequences and then apply the GMM estimator in Deaton and Laroque (1992) to those sequences. The properties of the collections of estimates for γ in all cases are reported in Table 2.1.

We see very clearly that ignoring the partial trend rather than the small-sample bias is the main source that the estimator tends to underestimate γ. This is evident by the fact that the 75 percentile values of the two cases with a partial trend are smaller than the true, meaning that at least 75 percent of the estimates are downward biased. In contrast, the empirical distribution of the estimates of γ when there is no trend has roughly equal weights on both sides of the true.

In line with the observations about the sample autocorrelation, this numerical analysis also shows that ignoring a mild partial trend can have significant effects on empirical results. This offers further support, besides the asymptotic result above, to the conjecture that the implausibly small estimates for the multiplicative factor in Deaton and Laroque (1992) are due to ignoring a downward trend starting at least from the year 1961.

2.4 The GMM Estimators

A GMM estimator of the trending storage model can work with three conditions: the autoregressive relation of detrended prices, and the trends in real prices and per capita productions. Following the conventional belief in the literature that global quantity data are unreliable, we only work with the first two conditions of real prices.

First, we have the autoregressive relation of detrended real prices:

$$u_{1,t} = \frac{p_t}{\lambda^t} - \frac{\gamma}{\lambda} \min \left\{ \zeta^*, \frac{p_{t-1}}{\lambda^{t-1}} \right\}. \tag{2.29}$$

Second, the process of real price contains an exponential trend:

$$p_t = \lambda^t \zeta_t.$$

Taking natural logarithm on both sides and rearranging yields:

$$\ln p_t = E_\infty \ln \zeta_t + t \ln \lambda + \left(\ln \zeta_t - E_\infty \ln \zeta_t \right), \tag{2.30}$$

where E_∞ stands for the mean according to the limit distribution of $\ln \zeta_t$.

Denoting $\alpha \equiv E_\infty \ln p_t$ and $u_{2,t} \equiv \ln \zeta_t - \alpha$, we have:

$$u_{2,t} = \ln p_t - \alpha - t \ln \lambda. \tag{2.31}$$

In the rest of this section, I introduce three GMM estimators: Estimator I, II and III. I and II recognize the price trend while III ignores it. I propose the use of I and the other two are presented for comparison.

Estimator I

Similar to the strategy of Eichenbaum and Hansen (1990), our first estimator works with the autoregressive relation of price and the trend in real price simultaneously.

Specifically, we estimate $(\gamma, \zeta^*, \lambda, \alpha)$ by minimizing $(v_I Z_I)' W_I (v_I Z_I)$ where:

$$v_I = \left[u_{1,2}, ..., u_{1,T}, u_{2,1}, \ldots, u_{2,T} \right]'; \tag{2.32}$$

$$Z_I = \begin{bmatrix} 1 & \min \{\xi, p_0\}^{-1} & I_{\{\xi \le p_0\}} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \min \left\{ \xi, \frac{p_{T-1}}{\lambda^{T-1}} \right\}^{-1} & I_{\left\{ \xi \le \frac{p_{T-1}}{\lambda^{T-1}} \right\}} & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{T} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & \frac{T-1}{T} \end{bmatrix}, \tag{2.33}$$

and W_I is the weighting matrix.[15] The parameter ξ is an initial guess of the cutoff price. I specifically defined it as

$$\xi \equiv Percentile \left(\left\{ \frac{p_t}{\lambda_{OLS}^t} \right\}_{t=0}^{T}, \mu \right), \tag{2.34}$$

[15] Variable t/T is used as instrument for trend estimation as in Eichenbaum and Hansen (1990).

the μ percentile value of the sample price. In the other words, the percentage μ is a rough guess of the stock-out rate of the sample price.

The specification of the second and third column of the instrument matrix roughly approximates the optimal instrument, with which a GMM estimator of the storage model attains the semi-parametric efficiency bound for the conditional restriction of price disturbance. Their resemblance of the functional properties of the optimal instrument is expected to bring some finite-sample efficiency gain.[16]

In practice, I choose $\mu = 0.85$. We can first use the OLS implied decay factor, λ_{OLS}, to construct the instrumental matrix Z_I and sub-optimal weighting matrix $\left(Z_I' Z_I\right)^{-1}$, with which we can obtain a vector of preliminary estimates. I then update the instrumental matrix by replacing ξ with the preliminary estimate of ζ^* and λ_{OLS} with the preliminary estimate of λ, and calculate the estimated optimal weighting matrix for the second round of estimation. I repeat such updating until the estimates become stable.

The Appendix will provide a robustness check to the value of μ and the number of iterations for this and the rest two estimators. There, we will see that for all three estimators the estimates starting either from $\mu = 0.85$ or 0.90 become the same after a few iterations. This suggests that all the three estimators are robust to the value of μ within the studied range conditional on a given number iterations is reached.[17]

Estimator II

The conventional empirical approach to trending data would first detrend the price data and then estimate the autoregressive relation of the detrended price using the previously detrended price data. With λ_{OLS} known from detrending, we can rewrite (2.29) as:

$$u_{1,t}^{II} = \frac{p_t}{\lambda_{OLS}^t} - \frac{\gamma}{\lambda_{OLS}} \min\left\{\zeta^*, \frac{p_{t-1}}{\lambda_{OLS}^{t-1}}\right\}. \tag{2.35}$$

We can estimate $(\gamma, \zeta*)$ by minimizing $(v_{II} Z_{II})' W_{II} (v_{II} Z_{II})$ where:

$$v_{II} = \left[u_{1,1}^{II}, ..., u_{1,T}^{II}\right]; \tag{2.36}$$

$$Z_{II} = \begin{bmatrix} 1 & \min\{\xi, p_0\}^{-1} & I_{\{\xi \le p_0\}} \\ \vdots & \vdots & \vdots \\ 1 & \min\left\{\xi, \frac{p_{T-1}}{\lambda^{T-1}}\right\}^{-1} & I_{\left\{\xi \le \frac{p_{T-1}}{\lambda^{T-1}}\right\}} \end{bmatrix}, \tag{2.37}$$

and W_{II} is the weighting matrix.

---

[16]For details, see Chapter 1 of this thesis. While Chapter 1 proposed a more sophisticated instrument, it also suggests that any function of lag-one price with functional properties that resemble those of the optimal instrument can be a good choice.

[17]The estimation of the variance-covariance matrix for this estimator should recognize the trend as well. We refer to Eichenbaum and Hansen (1990) and do not repeat here.

The parameter ξ is defined as in (2.34). In practice, to be consistent with Estimator I, I choose $\mu = 0.85$. And, similar to Estimator I, I first use λ_{OLS}, together with μ, to construct the instrument matrix and the sub-optimal weighting matrix, with which we can obtain preliminary estimates. I then update the instrument matrix and the implied optimal weighting matrix and estimate again. I iterate until the estimates become stable.

Note that, when updating using previous estimates, we only update the cutoff price in the instrument matrix and the weighting matrix. Unlike in Estimator I, we do not update the decay factor λ as it is always fixed at λ_{OLS} and never estimated again. This is the major difference between the Estimator I and II and an important feature of two-step estimators in general.[18][19]

While both Estimator I and II recognize trend, we should pay attention to their relative finite-sample performances. Since Estimator II remains a popular empirical strategy to the trending data while Estimator I is not common, a numerical study of the performances of the two at finite sample-size is interesting and useful.

Such comparison of empirical strategies is particularly important to the empirical literature of the storage model where only small sample data are available.[20] I offer a numerical comparison of Estimator I and II in the next section.

Estimator III

Assuming no trend, the auto-regressive relation of price (2.29) reduces to its classic form:

$$u_{1,t}^{III} = p_t - \gamma \min \{\zeta^*, p_{t-1}\}. \tag{2.38}$$

[18]Ignoring the first step in calculating the standard errors can lead to inconsistent standard errors for the second step. Theorem 6.2 of Newey and McFadden (1994) states that the first step estimator affects the second step standard errors if and only if inconsistency in the first step leads to inconsistency in the second step (page 2180, Newey and McFadden, 1994). In the current situation, if the first step estimator of the decay factor is inconsistent, we will not have consistent estimation of γ and ζ^* in the second step. Therefore, calculation of standard errors of the second step of the Estimator II must incorporate variance of the first step estimator (page 2183, Newey and McFadden, 1994).

[19]Under some conditions, the first step is super-consistent. However, it is unclear that if it is super-consistent in all importance cases of the storage model. For example, in the model considered in Bobenrieth, et al. (2008), the stationary price is unbounded and its expectation is approaching to infinity. When their model is the detrended counterpart of the trending storage model considered here, it is unclear if the intercept of the OLS detrending is infinite or not.

[20]Since high-frequency short-time-span data commonly seen in empirical finance contains much noise and does not include enough influential commodity booms, the empirical literature of storage model usually uses low-frequency long-time-span data such as annual average prices. Also, since structure changes like technical improvement in commodity markets are common, time relevant inference further restricts sample size. The sample size in Deaton and Laroque (1992, 1995 and 1996) is 88, in Cafiero et al. (2010) 89, and in Chambers and Bailey (1996) 384 (324 for coffee) but they have more parameters to estimate.

Then, we can estimate (γ, ζ^*) by minimizing $(v_{III}Z_{III})' W_{III} (v_{III}Z_{III})$ where:

$$v_{III} = \left[u_{1,1}^{III}, ..., u_{1,T}^{III}\right]; \tag{2.39}$$

$$Z_{III} = \begin{bmatrix} 1 & \min\{\xi, p_0\}^{-1} & I_{\{\xi \leq p_0\}} \\ \vdots & \vdots & \vdots \\ 1 & \min\{\xi, p_{T-1}\}^{-1} & I_{\{\xi \leq p_{T-1}\}} \end{bmatrix}, \tag{2.40}$$

and W_{III} is weighting matrix.

Since trend is assumed out, we have accordingly

$$\xi \equiv Percentile\left(\{p_t\}_{t=0}^{T},\ \mu\right). \tag{2.41}$$

Fixing μ at 0.85 again and using the weighting matrix $\left(Z'_{III}Z_{III}\right)^{-1}$, we can obtain preliminary estimates of ζ^* and a vector of residuals. As usual, we iterate to update the instrument matrix and the weighting matrix until the estimates reside. By ignoring the price trend, the calculation of the standard errors of the Estimator III is standard and wrong.

2.5 Empirical Result for the World Corn Market

The world annual average nominal price index for corn is from an updated version of the Pfaffenzeller, et al. (2007) data set, which is itself an update of the Grilli and Yang (1988) data set.[21] The real corn price is obtained by deflating the nominal price index using the MUV-G5 index, a manufacture unit value index of the G5 countries, from the same data set. The U.S. CPI is another popular deflator. If the U.S. CPI were used, the downward trend in the real price for corn would be stronger because the U.S. CPI grows much faster than the MUV-G5 in the recent decades. Therefore, switching to the U.S. CPI would mostly likely make the arguments in this chapter stronger.

I choose to use the sample period over 1961 - 2005. The green revolution was already widely applied at the year of 1961. The U.S. corn yield actually started to grow significantly since the 1940s, but still at a slower rate than after the 1961. At 2005, the projected corn use for ethanol experienced a dramatic increase from the previous projections for the year 2006 and beyond. This sudden shift in projection may shift the expectation of the forward-looking storers. Given all these considerations, the sample period 1961 -2005 is chosen for the following empirical exercise.

The estimation result is reported in Table 2.2. Figure 2.8 plots the real prices of corn together with the implied cutoff prices under all three specifications. If the current price is above the estimated cutoff price, a stock-out is predicted.

From Table 2.2, all three specifications imply seven stock-outs in the relevant period. But, the timing of the stockouts when trend is ignored is actually quite different from the

[21]The data set is available at http://www.stephan-pfaffenzeller.com/cpi.html.

Table 2.2: Estimation result of the trending storage model

Parameters:	Recognize Trend		No Trend
	I	II	III
γ	1.0186	1.0394	1.0084
	(0.0253)	(0.0146)	(0.0268)
ζ^*	1.9248	1. 9368	1.3683
	(0.1326)	(0.1045)	(0.0861)
λ	0.9753	0.9751	
	(0.0016)	(0.0022)	
α	0.4725	0.4763	
	(0.0479)	(0.0565)	
S.O.	7	7	7
J_T	2.5277	2.1511	0.1733
	(0.8881)	(0.8575)	(0.3228)

Note. Standard errors in parentheses.
In the final row, probability values in parentheses.
J_T stands for the over-identifying test value.

other two. From Figure 2.8, when no trend is assumed, all the stockouts are concentrated in the first 16 years while all later prices are far below the constant cutoff price. This immediately implies that the world corn market is perfectly safe beyond the 70s. This is strongly inconsistent with the actual stock-to-use ratios (without China), whose historical lowest level occurs during the mid-90s boom (see Figure 2.3).[22] We know that the theory of storage suggests that the stock-to-use ratio is negatively correlated with price. Even though this simple model cannot capture all market disturbances, it is still hard to believe that the historical lowest stock-to-use ratio and a commodity market that is far from stock-out could have coexisted at the same time.[23] Therefore, the empirical implications for

[22]In the current model, stock-out simply means there is no speculative inventory. The estimation result suggests that there were seven occasions within the relevant period that the global corn speculative stocks are zero. The reported stocks data, however, are never zero. There are a few possible explanations. First, the reported stocks data contain non-speculative part, for example, the essential or "working stocks" that may only response to the growth of the economy rather than the short-run dynamics of prices, and the public reserves that may not managed entirely through the market. Second, Bobenrieth, et al. (2008) showed that when the production distribution has an atom at zero and the price at zero consumption is infinite, the cutoff price is infinite and there will be no stockout at all. Under the truth of their model, estimating the cutoff price is meaningless. Nevertheless, even with infinite cutoff price the market demand curve can still be highly nonlinear around some given price level and the estimated cutoff price may capture this point. In a heuristic sense, the "cutoff price" estimated in this way is still useful. For details regarding this issue see Chapter 2 of this thesis.

[23]In fact, we can also expect in general that if no trend is assumed when there is in fact a downward trend, implied stock-outs will always concentrate in earlier years of the sample. Replication of Deaton and Laroque (1992) using the updated data shows that most implied stock-outs are in the first few decades in the last century.

stock-outs of assuming no trend seem to make little economic sense.

In contrast, the implied trending cutoff prices under specifications I and II when the trend is recognized both identify three clusters of stock-outs, each at one of the three well-known commodity booms in the second half of the last century. The implied cutoff prices under I and II are close as evident in Figure 2.8. This is consistent with the similar estimates of $\zeta^*$and λ under I and II in Table 2.2. We also observe that the difference between the two cutoff prices becomes smaller as year goes by. This is, of course, an immediate implication of the exponential downward trend. It is also interesting to observe that the 2004 price almost reached the cutoff price under both I and II, but the price dropped in 2005. If there were not a good harvest in the marketing year 2004/05 (see Figure 2.2), corn market would have easily entered stock-out in 2005 and would have even less preparation for the later dramatic biofuel expansions.

Comparison among the estimates of γ in Table 2.2 is illuminating. Particularly, when no trend is assumed, the estimate of γ is very close to one, consistent with the finding of Deaton and Laroque (1992). Usually, such small estimate will invoke discussions about the existence of convenience yield. However, as we see in section 3, it may be explained as well by that a (partial) downward trend in real price is ignored. Taking the trend away before/during estimating the auto-regression of price disturbance is expected to yield larger estimate.

This is easily confirmed in Table 2.2. The estimated γ is 1.0394 under II and 1.0186 if I is used. The implied real interest rate using II would be too high for risk-neutral speculators. In contrast, a real interest rate of 1.86% (neglecting deterioration rate) implied by the simultaneous approach is in the very acceptable range without resorting to the existence of a convenience yield.

The over-identifying statistics also give interesting results. While we fail to reject none of the specifications, the test result that ignoring trend seems to fit the data quite well should be trusted. This artifact should be due to the failure of the over-identifying test in detecting trend misspecifications.

Unfortunately, the large standard error of the estimate of γ makes above comparison less conclusive. At such small sample size, we simply cannot avoid such problem. Nevertheless, though I fail to fully resolve the problem with inferring the expected real rate of return, the proposed estimator succeeds in attaining smaller standard error of the previous methods using smaller sample size. This is indeed encouraging progress.

Implied sample autocorrelation coefficients in the corn price

Table 2.3 presents the sample autocorrelation coefficients for the real price for corn for cases when the trend is recognized and ignored, and for the sample period 1961 - 2005 and 1900 - 1987.

Applying (2.19) directly to the corn real price gives the first- and the second-order sample autocorrelation coefficients at 0.8785 and 0.7443, respectively. Applying (2.19) to the corn real price after it is detrended using the estimated λ of Estimator I yields first-

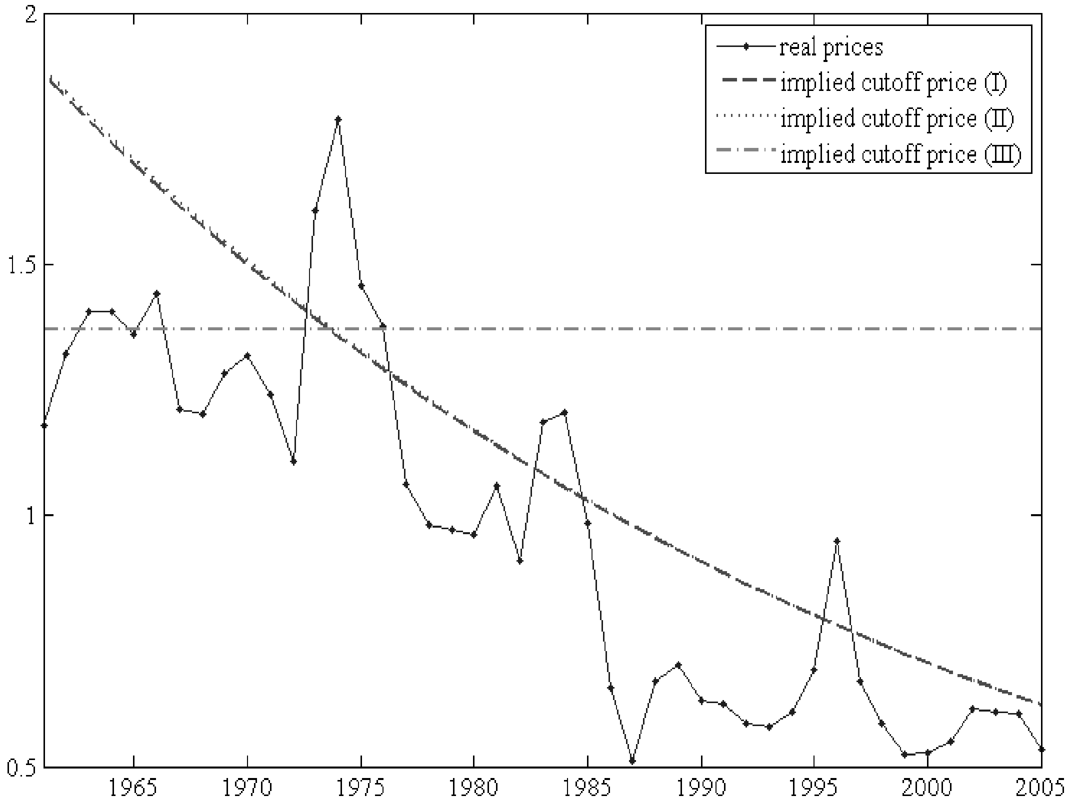

Figure 2.8: Implied cutoff prices

Table 2.3: Implied sample autocorrelation coefficients in the real price for corn

	AC(1)	AC(2)
1961 - 2005, trend over the entire sample		
ignore trend	0.8785	0.7443
recognize trend of I	0.5965	0.1758
recognize trend of II	0.5954	0.1735
upward bias	about 47%	about 300%
1900 - 1987, trend over sub-sample 1961 - 1987		
ignore trend	0.6488	0.3945
recognize trend of I	0.5253	0.1672
recognize trend of II	0.5255	0.1673
upward bias	about 23%	about 150%

Note. The AC(1) and AC(2) over 1900 - 1987 when trend is ignored is different from those in Deaton and Laroque (1992) because they deflated using U.S. CPI rather than MUV-G5.

and second-order sample autocorrelation coefficients at 0.5965 and 0.1758, respectively. If the estimated λ from Estimator II is used for detrending, the implied first- and second-order sample autocorrelation coefficients in price become 0.5954 and 0.1735. Both are smaller than if the trend is ignored as expected.

Noting that Deaton and Laroque's (1992, 1995, 1996) claim about the inability of their speculative model to generate the observed high autocorrelation in price is based on the data over the period 1900 - 1987, I conduct another check using their sample period assuming that the sub-sample from 1900 to 1960 is trendless and the rest follows the trend above estimated using the sample over 1961 - 2005.[24]

Applying (2.19) directly to the real corn price from 1900 to 1987 yields the first- and second-order sample autocorrelation coefficients at 0.6488 and 0.3945, respectively.[25] Detrending the sub-sample 1961 - 1987 using the estimated λ of Estimator I yields first- and second-order sample autocorrelation coefficients at 0.5253 and 0.1672, respectively. If the λ estimated using Estimator II is used, the first- and second-order sample autocorrelation coefficients become 0.5255 and 0.1673, respectively.

In sum, ignoring the price trend over the period 1961 - 2005 upward biases the first- and second-order sample autocorrelation coefficients by about 0.28 and 0.57, or 47% and

[24]This assumption is obviously ad hoc. For example, the US corn yield started to take off at 1940 though at a lower rate than after 1961. While 21 years of annual data from 1940 to 1960 is too small to accurately infer the trend over that period, taking into account of this lesser trend as well will likely enhance the current argument.

[25]Because this paper uses MUV for deflation whereas Deaton and Laroque (1992, 1995, and 1996) use U.S. CPI, the sample autocorrelation coefficients directly calculated from the data without detrending are different in the two.

Table 2.4: Small-sample performance of Estimator I and II

Model	Parameters	Percentiles	Estimator I	Estimator II
Low capital cost	$\gamma = 1.02$	25%	1.0162	1.0011
		75%	1.0594	1.0695
	$\zeta^* = 1.4074$	25%	1.1186	1.2103
		75%	1.5207	1.9086
High capital cost	$\gamma = 1.04$	25%	1.0277	1.0211
		75%	1.0863	1.1075
	$\zeta^* = 1.3485$	25%	1.0783	1.1933
		75%	1.5049	1.9478

Note: 25% and 75% stand for the 25 and 75 percentile values of the estimates.

3 times, respectively, and ignoring a partial trend in the price sample over the period 1900 - 1987 can still upward bias the first- and second-order sample autocorrelation coefficients by about 23% and 1.5 times, respectively. Such great distortion raises serious concerns on much of the previous comparisons between the observed and the simulated autocorrelation coefficients assuming no trend.[26]

1.86% or 3.94%? A numerical comparison between Estimator I and II

Using the same numerical method in section 3, I simulate 2000 sequences of 50-period trending prices and carry out a small numerical comparison between small-sample performances of Estimator I and II. The per capita inverse demand is assumed to be $f(c) = 1-c$, the population growth rate is 1%, the decay factor in price is 0.98, and the harvest is *i.i.d.* standard normal discretized into 20 nodes using the Gauss Hermite quadrature. Physical deterioration is ignored for simplicity. I experiment with two real interest rates: 2%, which is close to the estimation result of Estimator I, and 4%, close to that of II. The true ζ^* is 1.4074 for $r = 2\%$ and 1.3458 for $r = 4\%$.

Applying the two estimators to each sequence of the simulated prices, I produce two collections of estimates for each parameter. The statistical properties of the collections for γ and ζ^* are reported in Table 2.4.

It is easy to see that Estimator I performs better in estimating γ as evident by its closer 25 and 75 percentile values to the true. Though its 25 percentile values for ζ^* is too small, the 75 percentile value of II for ζ^* is way off. So, it would be reasonable to say that Estimator I is also better in estimating ζ^*. This exercise suggests that the Estimator II tends to have, in general, more biased estimate for γ and as a result provides some support to my previous conjecture that the estimate of γ using Estimator I is more reliable.

[26]The upward bias for the second-order autocorrelation coefficients are much larger than what we observe from the simulation analysis in section 3. This myth is left unsolved in this chapter. I conjecture that this may be related to the frequency of the corn price I used here.

Implied ex-post real rates of return

Using the estimates obtained above, we can calculate the sequence of the implied ex-post real rates of return using:

$$\left\{ \frac{p_t}{\min\left\{\hat{\zeta}^* \hat{\lambda}^{t-1},\ p_{t-1}\right\}} \right\}_{t=1}^{T}, \tag{2.42}$$

where $\hat{\lambda}$ and $\hat{\zeta}^*$ are estimates of Estimator I or II (see Figure 2.9).

Because the cutoff prices implied by I and II are very close, the sequences of ex-post real rates of return are also close as expected and so are their time averages (denoted by the horizontal lines in Figure 2.9). The simple time average of the ex-post real rates of return implied by I is 1.0160 and by II is 1.0155, both of which are much closer to the estimated γ using I than using II.

The discrepancy between these averages and the estimates of γ should not be surprising: the procedure to obtain those estimates involves implicitly weighting the residuals using the instruments and explicitly weighting the sample moments using the weighting matrix, whereas the calculation of the simple time averages here is naive. At large sample size, those averages should be closer to the GMM estimates of γ.

2.6 Conclusion

This chapter analyzes and estimates a model of rational expectations competitive storage with a price trend, which implies a serial correlation in price. I study the short-run dynamics of the price implied by this model particularly focusing on the first- and second-order autocorrelation coefficients and the expected rate of return. I present the empirical results of three GMM estimators and discuss their implications. This work provides further support for the theory of storage and a new perspective on the price behavior in commodity markets. The comparison and discussion of the empirical strategies is interesting from the methodological point of view, as well as practically relevant.

Besides the relatively large standard error of the estimated multiplicative factor, another limitation of this paper is that the production response to price is ignored. While introducing responsive production is theoretically possible (see Scheinkman and Schechtman, 1983; Wright and Williams, 1982, 1984), introducing additional parameters in estimation at such a small sample size would be practically difficult. The effects of this over-simplification, however, could be mitigated by the fact that the estimated short run production elasticity to price may be small.

A third limitation is the neglect of income growth, though it would be difficult to believe that the consumption of grains increases at the same rate of income. Indeed, the income elasticities of demand for grains are likely to be lower than for petroleum. We could follow a strategy similar to Dvir and Rogoff (2010) by assuming that the aggregate production is homogeneous to an exponentially decreasing proportion of the world

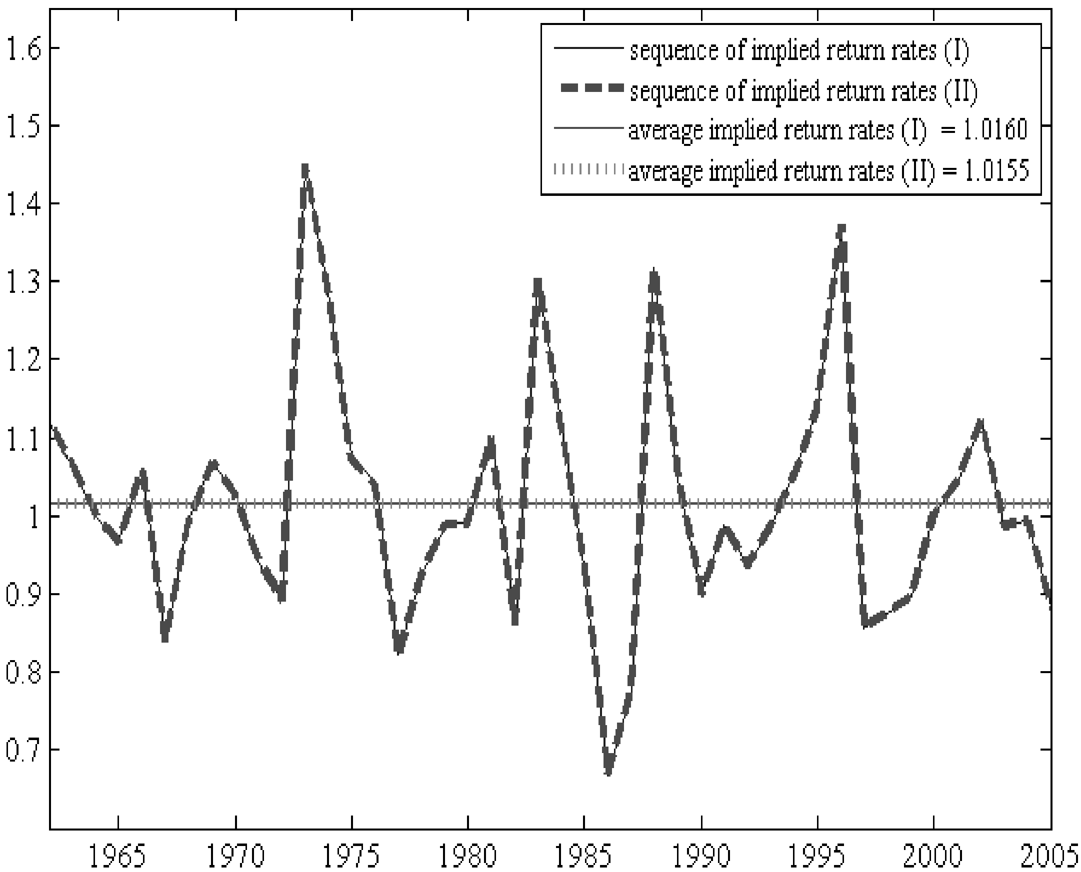

Figure 2.9: Implied ex-post real rates of return

real GDP. Nevertheless, without any affirmative theoretical or empirical support to this homogeneity assumption, we refrained from doing so. What has to be kept mind is that estimation using only price would be the same in either case.

Finally, the substitution effects among the major crops are ignored. The only well-founded multiple commodity storage model was developed by Nishimura and Stachurski (2009). Estimation of it is necessarily challenging.

Chapter 3

The Empirical Implications of Ignoring the Probability of Extreme Events: The Small-sample Performance of two Misspecified NLS Estimators of a Commodity Storage Model

3.1 Introduction

Proponents of limited-information estimators emphasize their robustness to distributional specifications that are not necessarily clear to the applied econometricians and thus must be assumed heuristically in the full-information estimators. Yet, the distributional robustness of the limited-information estimators is limited: some distributional changes may result in structural changes in the conditional restrictions on which those estimators are based. This chapter addresses the performance and empirical implications of limited-information estimation beyond the robustness to distributional specifications. In the context of commodity storage with a small probability that excess supply is zero, I show that misspecified estimators that ignore this small-chance event can perform better at finite-sample size. This is by no means a defense of the misspecification but a strong signal that in finite samples such misspecification can be hard to identify and proper interpretations of empirical results that recognize the potential of such misspecification would be necessary.

Our working model is a modified version of the commodity storage model in the tradition of Gustafson (1958). Its conventional counterpart satisfactorily explains important features of important features of commodity prices like skewness and infrequent but violent spikes (Williams and Wright, 1991; Deaton and Laroque, 1992). Estimation of the classic version was pioneered by Deaton and Laroque (1992) using a GMM estimator. This approach can estimate two important parameters: the real interest rate implied in the commodity market and the cutoff price, a price level above which the speculative stocks is depleted and the (current) backwardation occurs. It is usually believed that this approach is robust to specifications of the distribution of harvest shock and the demand curve.

Nevertheless, its robustness fails on potentially important dimension: the existence of extremely infrequent, extremely low harvest shock. With unbounded price and an atom at zero in the distribution of the harvest shock, the storage model implies that the commodity is continuously stored or equivalently the cutoff price above which stock-out occurs is at infinity. This type of model is studied in Bobenrieth, et al., (2002, 2004, 2008, and 2012). In contrast, the classic storage model where harvest shock is bounded away from 0 and/or price is bounded from above implies a finite cutoff price which can be passed a couple times within limited periods. The real interest rate, however, enters the auto-regression of price of both models.

Under the classic model, the cutoff price can be estimated using limited-information estimators whereas in the modified model estimating it is meaningless. The tiny chance of the occurrence of the extreme events renders their observation difficult within finite periods and thus easily ignored by the applied econometricians. But their extreme impacts can make the implied price behavior systematically different at far horizons. To understand the welfare and policy effects of such potential misspecification, we must understand how it effects the estimates based on each model as well as what the “wrong” estimates are capturing if they capture anything at all.

Assuming that the modified model is the true model, I numerically study the distribution of the estimators with and without recognition of the small probability at zero harvest. I find that, surprisingly, a misspecified estimator better estimates the real rate of interest and the (wrongly) estimated cutoff price actually captures the sharp turn of the equilibrium price function. The pessimistic implication is that just by observing finite price sample, it would be difficult to distinguish the two models though their economic implications on price behavior are dramatically different at far horizons. The somewhat comforting implication is that because the sharpest turning point represents the point that the conditional variance of forward price disturbance dramatically changes, information of this “wrong” cutoff price turns out to be useful.

It is important to clarify that misspecification in this chapter is a relative term. It is unclear that as the sample size increase whether the estimator for the “wrong” cutoff price will go to infinity, the true cutoff price, or will stay around the sharp turning point of the equilibrium price function.[1] How the two estimators studied in this chapter behave at large sample size is of course an important and interesting question to ask in the future work.

The rest of the chapter is organized as follows. Section 2 introduces the modified commodity storage model. Section 3 numerically studies the performance and empirical implications of the two misspecified estimator, and section 4 concludes.

3.2 A Commodity Storage Model with an Atom at Zero Harvest

Consider a competitive market for a single storable commodity. Time is discrete. All agents have rational expectations.

The harvest is assumed as an exogenous *i.i.d.* disturbance $\omega \in [0, \bar{\omega}]$, $0 < \bar{\omega} < \infty$, and ω has a mixed discrete-continuous distribution with a countable set of atoms, one of which is at zero. More precisely, the distribution of ω is of the form

$$\alpha L_d + (1 - \alpha) L_c, \tag{3.1}$$

where $\alpha \in (0, 1)$, L_d is a discrete distribution that has an atom at 0, and L_c is an absolutely continuous distribution, with continuous derivative when restricted to its support $[0, \bar{\omega}]$.

Assume that there is a continuum of identical producers, a continuum of identical storers, and a continuum of identical consumers; each of the three has total measure of one. Let ω' be the next period's harvest. Storers can hold output from one period to the next, and the sole cost of storage is the cost of of capital invested. Given storage x, the next period's total available supply is $z' \equiv x + \omega'$.

Producers and storers are risk neutral and have a common constant discount factor $\delta \equiv 1/(1 + r)$, where $r > 0$ is the discount rate.

[1]The author thanks Michael Jansson to point this out.

The utility function of the representative consumer $U : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous, once continuously differentiable, strictly increasing and strictly concave. It satisfies $U(0) = 0$, $U^{'}(0) = \infty$. The inverse consumption demand curve is then $f = U^{'}$. We assume U has finite upper bound, and thus total revenue $cf(c)$ is also bounded.

The perfectly competitive market yields the same solution as the surplus maximization problem. The Bellman equation for the surplus problem is:

$$v(z) = \max_x \left\{ U(z-x) + \delta E\left[v\left(z^{'}\right)\right]\right\}, \quad \text{subject to} \tag{3.2}$$

$$z^{'} = x + \omega^{'},$$

$$x \geq 0, \; z - x \geq 0,$$

where $E[\cdot]$ denotes the expectation with respect to next period's harvest $\omega^{'}$.

By standard results (see for example Stockey and Lucas with Prescott, 1989), v is continuous, strictly increasing, strictly concave, and the optimal policy function $x(z)$ is single valued and continuous.

Consumption and price are given by the function $c(z) \equiv z - x(z)$, $p(z) \equiv f(z - x(z))$.

The policy function x satisfies the Euler conditions:

$$f(z - x(z)) \geq \delta E\left[v^{'}\left(x(z) + \omega^{'}\right)\right], \quad \text{with equality if } x(z) > 0, \tag{3.3}$$

and the envelope condition $v^{'}(z) = f(z - x(z))$.

Given initial available supply $z > 0$, condition (3.3) implies that $z^{'} > 0$ and $x\left(z^{'}\right) > 0$, and this arbitrage condition holds with equality in the current period and for the indefinite future.

Letting $\gamma = \delta^{-1}$,

$$E_t p_{t+1} = \gamma p_t. \tag{3.4}$$

Since stocks are always positive, there is no stock-out and the auto-regression of price is:

$$p_{t+1} = \gamma p_t + u_{t+1}, \qquad \forall t \geq 0, \tag{3.5}$$

where u_{t+1} is innovation, i.e., $E_t u_{t+1} = 0, \; \forall t \geq 0$.

In contrast, in the classic storage model,

$$E_t p_{t+1} = \gamma \min\left\{p^*, p_t\right\}, \tag{3.6}$$

where p^*, the cutoff price, is defined as

$$p^* = \gamma^{-1} E_t p_{t+1} \text{ when } x_t = 0. \tag{3.7}$$

That is, when the storage at period t is zero, the discounted expected forward price is a constant and depends only on the forward harvest and the equilibrium price function. The cutoff price p^* is the minimum price such that stock-out occurs. In the other words, for all current prices that are higher than p^*, the stocks are zero and the discounted expected forward price is p^*.

Therefore, the auto-regressive relation of the classic storage model is:

$$p_{t+1} = \gamma \min\{p_t, p^*\} + u_{t+1}, \qquad \forall t \geq 0. \tag{3.8}$$

It is important to realize that in the modified, p^* can be understood as at infinity.

In the following analysis, I assume that the modified model is the true. For convenience, the modified model is called the atom model and the classic model the non-atom model.

3.3 Simulation Analysis

3.3.1 Generation of Simulated Prices

To generate sequences of price under the atom model, we first need to solve the atom model. The solution algorithm is complicated by the fact that the price function and the price are unbounded under the truth of the model. To avoid this difficulty, I approximate the price function only starting from a extremely small positive number in the state space, i.e., the availability. And, when using the backward solution algorithm to solve for the second to last equilibrium price function, I shift slightly the last period equilibrium price function, i.e., the inverse demand function itself, to avoid facing infinite forward price. This shift will not alter the convergence to the stationary equilibrium as the fixed point theorem predicts that the equilibrium price functions should converge to the unique stationary rational expectations equilibrium starting from any initial price function.

For the rest of the chapter, the following parameterization is used: the inverse demand function is $f(c) = c^{-2}$; the real interest rate is 0.04 and the deterioration rate is ignored for simplicity; the harvest shock is the log-normal with mean 1 and standard deviation 1 with 1% of the mass redistributed to the origin.

Figure 3.1 shows the equilibrium price function in log-log space and linear-log space. We can see while there is always storage as evident by the fact that the equilibrium price and the inverse demand never merge, the distance between the two becomes very small at relatively high price levels.

Figure 3.1 also gives the curvature of the equilibrium price function under two formulas specified in the figure. It is important to observe that the equilibrium price function has a sharp turning point around which the shrinking rate of the speculative storage dramatically decreases. This property gives the impression that the equilibrium price function has a kink just like the equilibrium price function of the non-atom storage model.

Figure 3.2 presents the distribution of the logarithm of the simulated prices. As we can see, there are very rare but extremely violent jumps in the price series. Figure 3.3 shows that the distribution of the corresponding simulated total availabilities. Consistent with huge price jumps, there is a very small chance that the availability hits extremely low levels.

3.3.2 The NLS Estimators

We consider two nonlinear least squares (NLS) estimators. The Level estimator which works directly with the price disturbance:

$$u_{t+1} = p_{t+1} - \gamma \min \{p^*, p_t\}, \quad \forall t \geq 0. \tag{3.9}$$

The Ratio estimator which works instead with the auto-regression of price change ratios:

$$v_{t+1} = \frac{p_{t+1}}{p_t} - \gamma \min \left\{ \frac{p^*}{p_t}, 1 \right\}, \quad \forall t \geq 0. \tag{3.10}$$

In the atom model, $p^* = \infty$ and then the minimum structure in the above two auto-regressive relations reduces to γp_t. Since we have assumed that the atom model is the true, the applied econometrician who assumes the false model will attempt to estimate the cutoff price p^*, which in principle should be at infinity. The estimates under either case are obtained by minimizing the sum of squared errors above. The level estimator for the parameters under the misspecification are denoted as $p^*_{L,M}$ and $\gamma_{L,M}$ while under the correct specification the estimator for γ is denoted as $\gamma_{L,C}$. The ratio estimator for the model parameters under the misspecification are denoted as $p^*_{R,M}$ and $\gamma_{R,M}$ while under the correct specification the estimator for γ is denoted as $\gamma_{R,C}$.

3.3.3 Estimation of the Inverse Discount Rate

I study the distribution of estimates of the NLS estimators using Monte Carlo Simulation by estimating each sequence using the two estimators under both misspecified and correctly specified estimators. By doing so, I can obtain approximate distributions of estimators for both gamma and the cutoff price.

I first study the estimation of the inverse discount rate. Table 1 reports the performance of both Level and Ratio estimators with misspecification. In this and the following simulations, the parameterization is the same as for the numerical examples in the last section unless specially notified. The total number of replications for each experiment is 10000 and sample size for each replication is always 100, small but about the size of reliable world annual average commodity prices. In each table, 25%, 50% and 75% stand for the 25, 50 and 75 percentile values of each collection of estimates, STD stands for the standard deviation of each collection and RMSE, the root mean square error, a statistic for the average deviations of the estimates from the true. For p^*, the RMSE is not calculated.

It is easy to see that the Ratio estimator is much better in estimating the inverse discount rate. The RMSE square error is 32.18% smaller than that of the Level estimator.

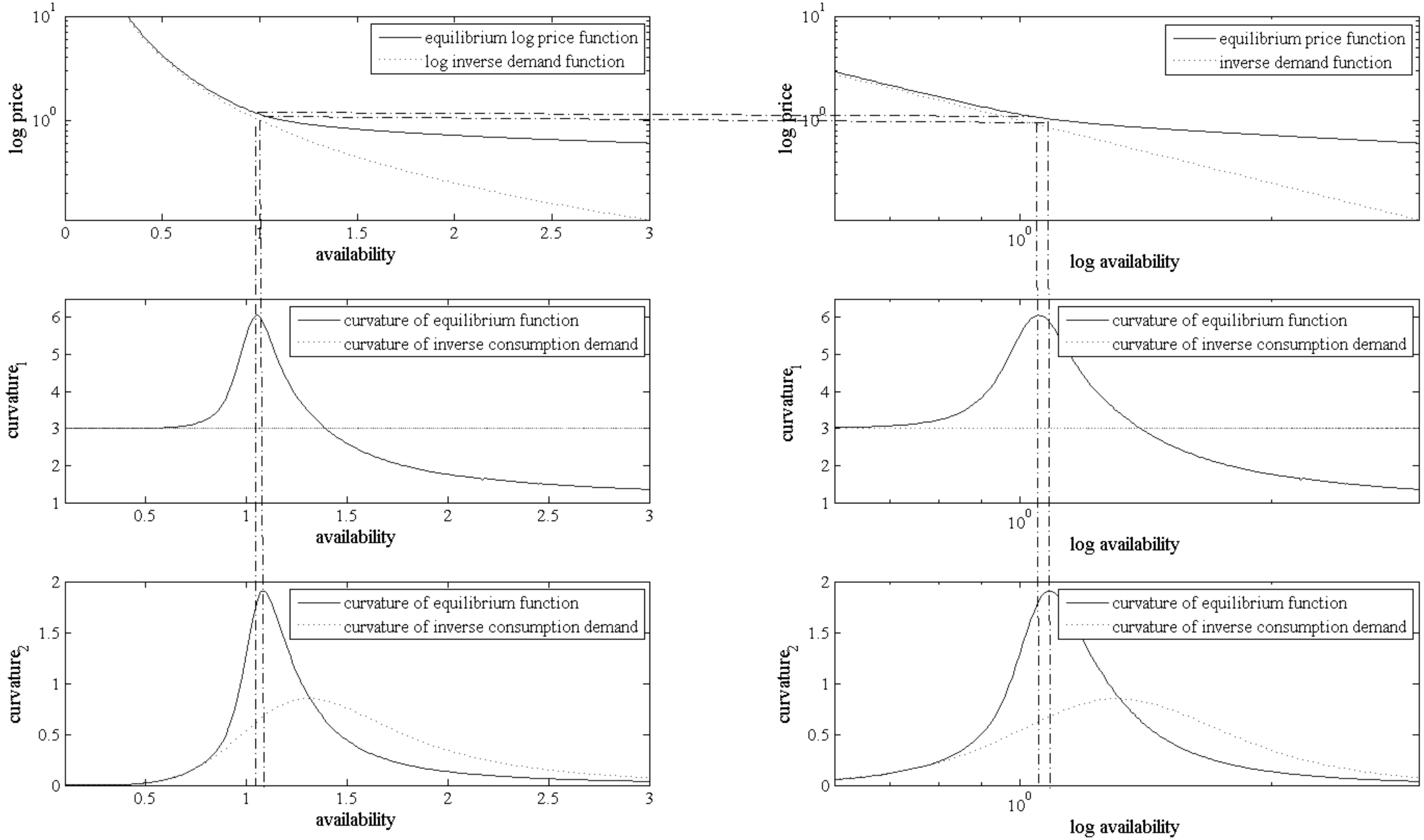

Figure 3.1: Curvature 1 for a function $y = f(x)$ is defined as: $\frac{-xf''(x)}{f'(x)}$ and curvature 2 as: $\frac{|f''(x)|}{\left(1+(f'(x))^2\right)^{1.5}}$.

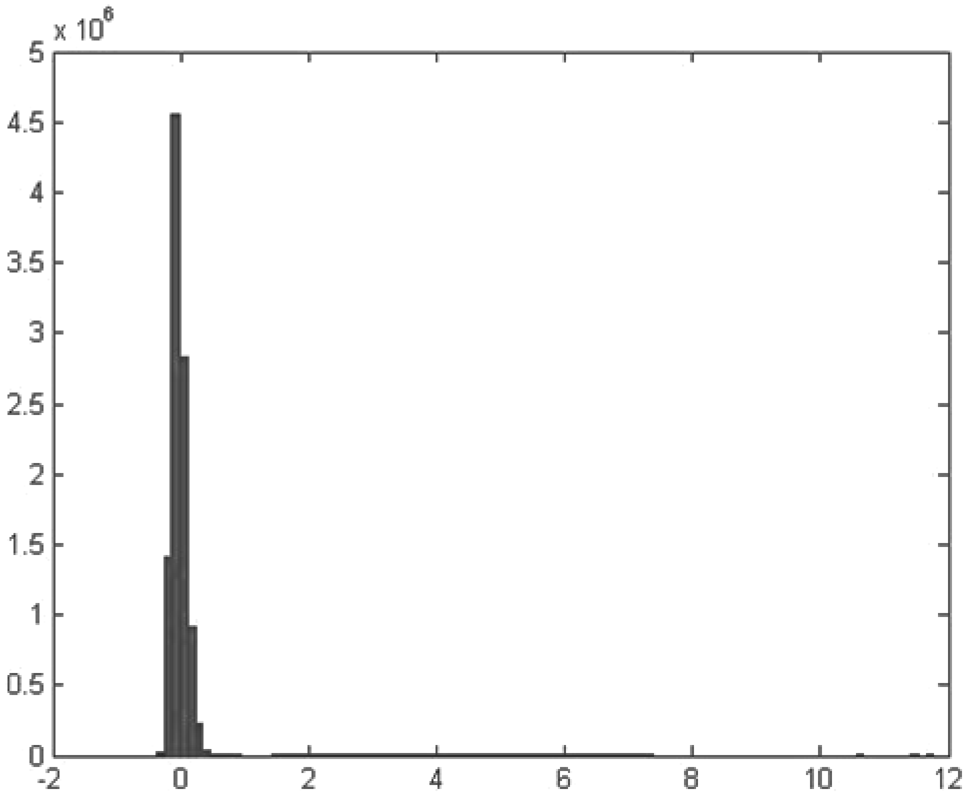

Figure 3.2: Distribution of logarithm of simulated prices

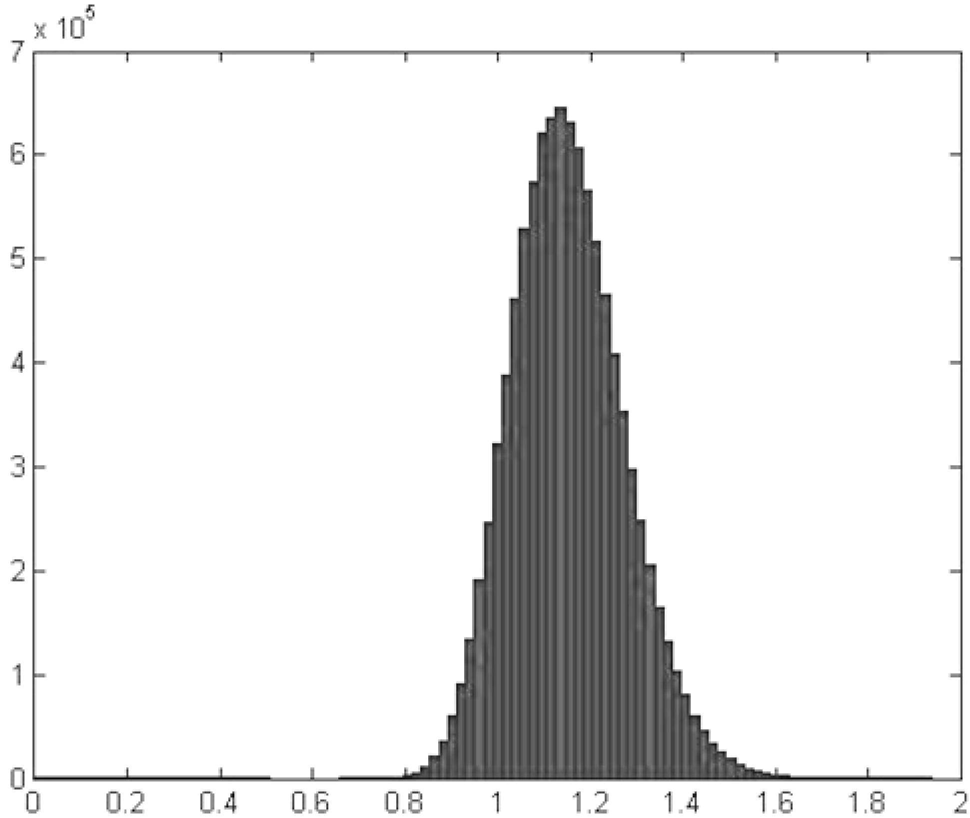

Figure 3.3: Distribution of simulated availabilities

The mean, 25, 50, and 75 percentiles values of the Ratio estimator are also closer to the true. One possible reason for this is that the conditional heteroskedasticity of the process of price-change-ratio disturbance is less strong than that of the price disturbance. The disturbance of the Ratio estimator is equivalent to the disturbance of the Level estimator re-weighted using the inverse lag-one price. We know that the conditional variance of price disturbance is increasing in the lag-one price, so dividing by the lag-one price down-weights the disturbance with high conditional variance. Therefore, the Ratio estimator is a variance stabilized Level estimator. Chapter 2 of this thesis studies the performance of inverse lag-one price as instrument under the GMM framework. A similar result relative to constant instrument is obtained.

Table 2 reports the estimation result under the correction specification. In this experiment, both estimators only estimate γ and recognize that p^* is infinite. The RMSE of the Level estimator is much smaller than that of the Ratio estimator. But this should not be understood as an evidence for the superiority of the Level estimator of the Ratio one. A closer scrutiny to the percentile values reveal that the Ratio estimator actually yields more estimates that are closer to the true. The larger RMSE and STD of the Ratio estimator may be due to a few extreme outlier estimates. The existence of these extreme outliers offsets the advantage of the Ratio estimator that more of its outcomes are closer to the true.

Comparison with the result in Table 1 yield, surprisingly, that the estimation of the inverse discount rate is better under the misspecification. The percentile values of the both the Level and the Ratio estimators with misspecification are closer to the true than without misspecification. This seems to suggest that it could be better to use misspecified model either intentionally or unintentionally regarding estimating γ.

One possible explanation is that putting a p^* in the disturbance ignores the case where price increases in expectation at the inverse discount rate starting from very high prices. Despite its validity, testing this implication of the atom model at small sample is difficult. The chance of a further price increase starting from extremely high price almost cannot be observed within finite periods. Therefore, most price change starting from extremely high price is negative and thus significantly drags down the estimate of γ. As a result, restricting the effects of those big dives by lowering the its starting value maybe helpful in estimation of γ.

3.3.4 Distribution of Misspecified Estimator of p^*

The distribution of the estimates for p^* and their relation to the curvature of the price function are presented in Figure 3.4 and Figure 3.5 under two formulas of curvature.

We see that the range from 25% to 75% of the estimator distribution matches the peak of the curvature of the price function. It seems that, though the econometric model is misspecified and estimating the p^* is meaningless in principle, the estimates seem to represent the sharp turn of the equilibrium price function.

This should not be surprising if it were combined with the result in 3.1 aforementioned.

Table 3.1: The small-sample performance: misspecified

	$\gamma = 1.04$	$p_0^* = \infty$	$\gamma = 1.04$	$p_0^* = \infty$
	$\gamma_{L,M}$	$p_{L,M}^*$	$\gamma_{R,M}$	$p_{R,M}^*$
Mean	1.0682	1.0328	1.0620	1.0263
25%	1.0149	0.9962	1.0173	0.9910
Median	1.0218	1.0247	1.0236	1.0191
75%	1.0297	1.0595	1.0310	1.0524
STD	0.3703	0.0617	0.2508	0.0588
RMSE	0.3713	/	0.2518	/

Note. 10000 replications with sample size 100.

Table 3.2: The small-sample performance: correctly specified

	$\gamma = 1.04$	$p_0^* = \infty$	$\gamma = 1.04$	$p_0^* = \infty$
	$\gamma_{L,C}$		$\gamma_{R,C}$	
Mean	0.9581	/	1.0382	/
25%	0.9938	/	1.0029	/
Median	0.9953	/	1.0040	/
75%	0.9966	/	1.0053	/
STD	0.1770	/	0.2423	/
RMSE	0.1950	/	0.2423	/

Note. 10000 replications with sample size 100.

The sharp turn on the equilibrium price function of the atom model resembles the kink on the equilibrium price function of the non-atom model, and *vice versa*. Under the truth of the atom model, the misspecified estimator can be heuristically understood as abstracting the sharp turning region into a single point. This explanation is sound particularly because the limited information estimator does not distinguish the inverse demand function and the equilibrium price function.

This observation is important for two reasons: first, the pessimistic implication is that by just observing a price sample, it would be difficult to distinguish the two models though their economic properties are systematically different at far horizon; and second, the somewhat comforting implication is that because the sharpest turning point represents the point that the conditional variance of forward price disturbance dramatically changes, information of this "wrong" cutoff price is in fact practically useful.

3.4 Conclusion

This chapter shows that misspecification can be a relative concept regarding estimation performance. The misspecified estimator of the atom model can be more relevant and useful at small sample size than the correctly specified estimator. Such observation should generalize to cases where the cutoff price is finite but extremely high.

This, however, should not be understood as a defense of ignoring the infrequent extreme events in the empirical analysis in asset pricing problems. That the misspecified estimators behaves as if it is abstracting the sharp turning region into a single point is an explanation for the interesting observation above but never a solution to the difficulty of empirically testing the atom model at finite-sample size. Future work on this is obviously warranted.

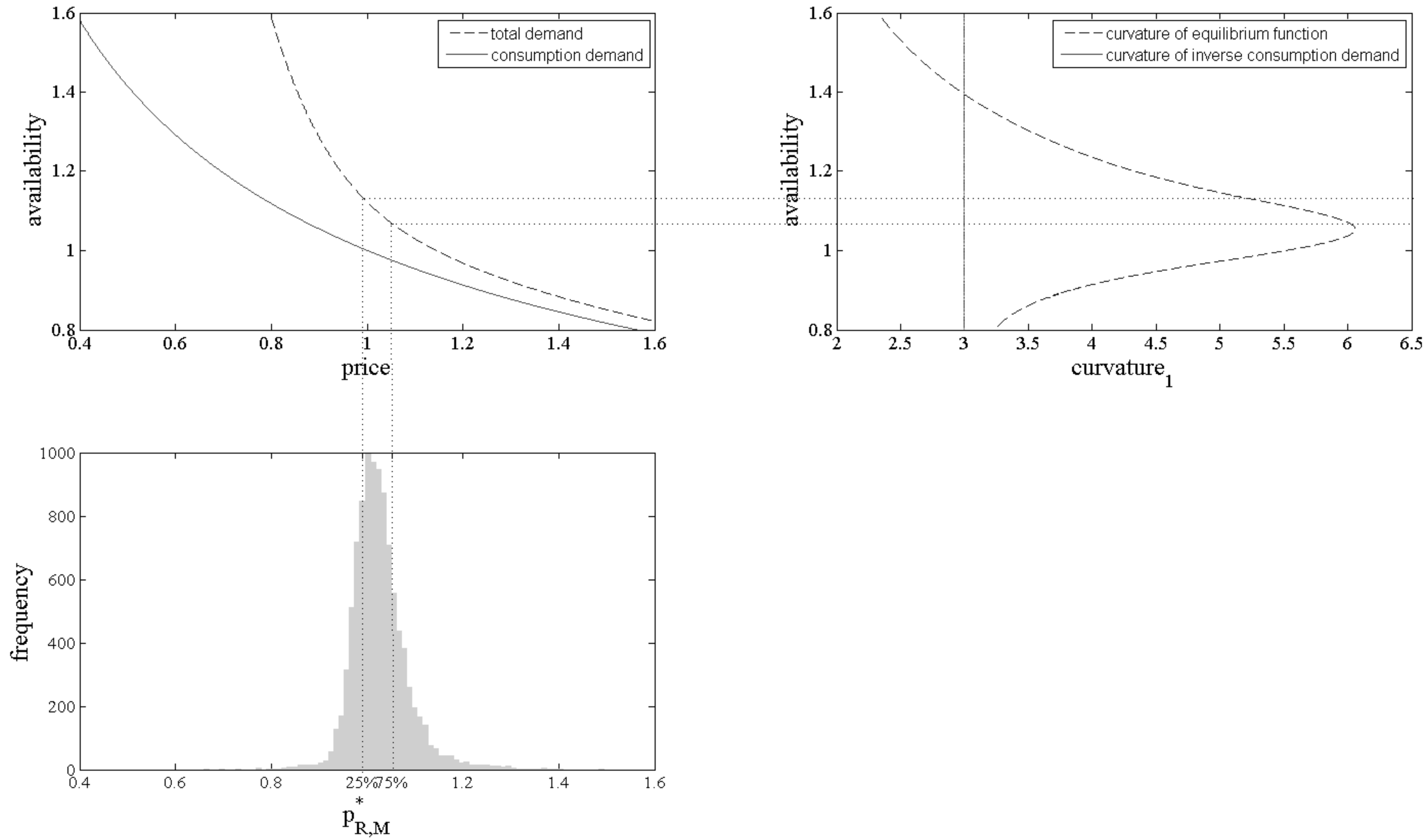

Figure 3.4: Curvature for a function $y = f(x)$ is defined as: $\frac{-xf''(x)}{f'(x)}$. The 25% and 75% lines stand for the 25 and 75 percentile values of the estimates.

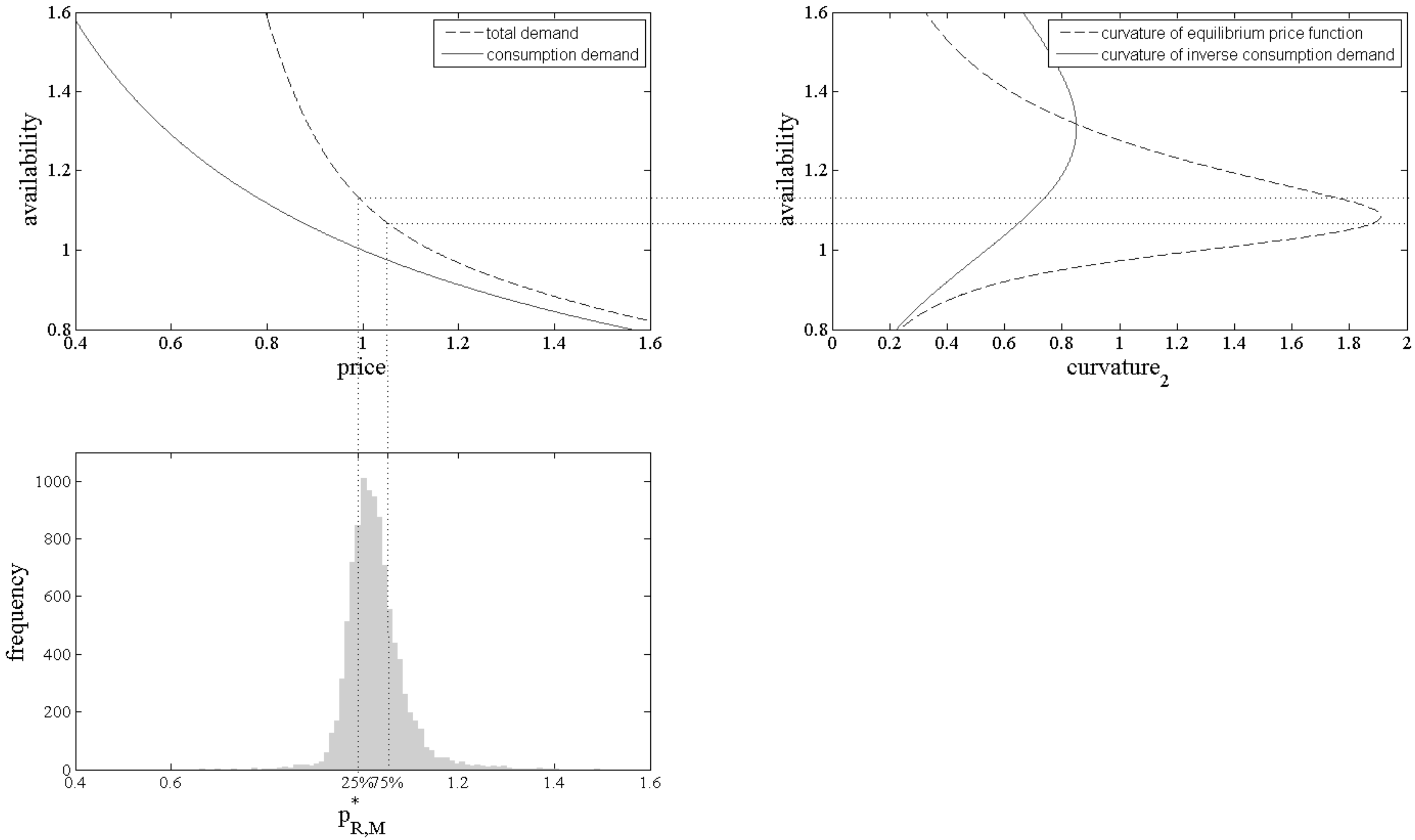

Figure 3.5: Curvature for a function $y = f(x)$ is defined as: $\frac{|f''(x)|}{\left(1+(f'(x))^2\right)^{1.5}}$. The 25% and 75% lines stand for the 25 and 75 percentile values of the estimates.

Bibliography

[1] Altonji, J. G. and L. M. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14(3), 353-366.

[2] Amemiya, T. (1977): "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955,968.

[3] Anatolyev, S. (2007): "Optimal Instruments in Time Series: A Survey," Journal of Economic Surveys, 21(1), 143-173.

[4] Angrist, J. D., G. W. Imbens and A. B. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57-67.

[5] Angrist, J. D. and A. B. Krueger (1995): "Split Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business & Economic Statistics*, 13(2), 225-323.

[6] Bobenrieth E. S.A., J. R. A. Bobenrieth and B. D. Wright (2002): "A Commodity Price Process With a Unique Continuous Invariant Distribution Having Infinite Mean," *Econometrica*, 70, 1213-1219.

[7] ———— (2004): "Characterization of the Equilibrium Price Function for a Family of Storage Models with Unbounded Price Realizations," Working Paper, Universidad de Concepción, Universidad del Bío-Bío.

[8] ———— (2008): "A foundation for the solution of consumption-saving behavior with a borrowing constraint and unbounded marginal utility," *Journal of Economic Dynamics and Control*, 32(3), 695-708.

[9] Bobenrieth E. S.A. and J. R. A. Bobenrieth (2010): "Storage Model with Exponential Price Trend," Unpublished Manuscript.

[10] Brennan, D., J. Williams and B. D. Wright (1997): "Convenience yield without the convenience: a spatial–temporal interpretation of storage under backwardation," *The Economic Journal*, 107, 1009-1022.

[11] Breusch, T., H. Qian, P. Schmidt and D. Wyhowski (1999): "Redundancy of Moment Conditions," J*ournal of Econometrics*, 91(1), 89-111.

[12] Broze, L., C. Francq, J.-M. Zakoïan (2001): "Non-redundancy of high order moment conditions for efficient GMM estimation of weak AR processes," *Economics Letters*, 71, 317–322. Buse, A. (1992): "The Bias of Instrumental Variable Estimators," Econometrica, 60, 173-180.

[13] Cafiero, C., E. Bobenrieth, J. Bobenrieth and B.D. Wright (2010): "A Maximum Likelihood Estimator of the Commodity Storage Model with an Application to the World Sugar Market," Working Papers 17-2010, Departamento de Economía, Universidad de Concepción.

[14] Cafiero, C., E. S. A., Bobenrieth, J. R. A. Bobenrieth and B. D. Wright (2011): "The empirical relevance of the competitive storage model," *Journal of Econometrics*, 162(1), 44-54.

[15] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," Journal of Econometrics, 34, 305-334. Chambers, M. J. and R. E. Bailey, (1996): "A Theory of Commodity Price Fluctuations," *Journal of Political Economy*, 104(5), 924-957.

[16] Chambers, M. J. and R. E. Bailey, (1996): "A Theory of Commodity Price Fluctuations," *Journal of Political Economy*, 104(5), 924-957.

[17] Donald, S. G. and W. K. Newey (2000): "A Jackknife Interpretation of the Continuous Updating Estimator," *Economic Letters*, 67, 239-243.

[18] Davidson, R. and MacKinnon, J. G. (1993): *Estimation and Inference in Econometrics*, Oxford University Press, New York.

[19] Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1221-1248.

[20] Deaton, A., and G. Laroque (1992): "On the Behavior of Commodity Prices," *Review of Economic Studies*, 59(1), 1-23.

[21] ———— (1995): "Estimating a Nonlinear Rational Expectations Commodity Price Model with Unobservable State Variables," *Journal of Applied Econometrics*, 10(S), S9-S40.

[22] ———— (1996): "Competitive Storage and Commodity Price Dynamics," *Journal of Political Economy*, 104(5), 896-923.

[23] ————(2003): "A Model of Commodity Prices after Sir Arthur Lewis," *Journal of Development Economics*, 71(2), 289-310.

[24] Dvir, E., and K. S. Rogoff (2010): "The Three Epochs of Oil," working paper, Boston College.

[25] Eichenbaum, M.S. and L. P. Hansen (1990): "Estimating models with intertemporal substitution using aggregate time series data," *Journal of Business & Economic Statistics*, 8 (1), 53-69.

[26] Ferson, W. and S. Foerster (1994): "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Model," *Journal of Financial Economics*, 36, 29-55.

[27] Fisher, R. A. (1925): *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

[28] Gallant, A. and G. Tauchen (1996): "Which Moments to Match?" *Econometric Theory*, 12(4), 657-681.

[29] Gourieroux, C., A. Monfort and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85-S118.

[30] Grilli, E. and M. C. Yang (1988): "Primary Commodity Prices, Manufactured Goods Prices, and the Terms of Trade of Developing Countries: What the Long Run Shows," *The World Bank Economic Review*, 2(1), 1-47.

[31] Gustafson, R. L. (1958): "Carryover Levels for Grains," USDA Technical Bulletin 1178, Washington, DC: United States Department of Agriculture.

[32] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029-1056.

[33] ———— (1985): "A Method of Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators", *Journal of Econometrics*, 30, 203-238.

[34] Hansen, L.P., J.C. Heaton and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83, 863-871.

[35] Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14(3), 262-280.

[36] Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectation Models," *Econometrica*, 50(5), 1269-1286.

[37] Harvey, D. I., N. M. Kellard, J. B. Madsen, and M. E. Wohar (2010): "The Prebisch-Singer hypothesis: four centuries of evidence," *The Review of Economics and Statistics*, 92, 367-377.

[38] Kim, Y., H. Qian and P. Schmidt (1999): "Efficient GMM and MD Estimation of Autoregressive Models," Economics Letters, 62(3), 265-270.

[39] McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995-1026.

[40] Michaelides, A. and S. Ng (2000): "Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of three simulation estimators," *Journal of Econometrics*, 96, 231-266.

[41] Newey, W. K. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in Maddala, G. S., C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*, Vol. 11, Elsevier Science.

[42] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, vol. 4, Robert Engle and Daniel McFadden editors, Elsevier – North Holland.

[43] Ng, S. and F. J. Ruge-Murcia (2000): "Explaining the Persistence of Commodity Prices," *Computational Economics*, 16 (1/2), 149-171.

[44] Nishimura, K. and J. Stachurski (2009): "Equilibrium storage with multiple commodities," *Journal of Mathematical Economics*, 45, 80-96.

[45] Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

[46] Pagan, A. and A. Ullah (1999): *Nonparametric Econometrics*, Cambridge University Press.

[47] Pfaffenzeller, S., P. Newbold and A. Rayner (2007): "A Short Note on Updating the Grilli and Yang Commodity Price Index," *World Bank Economic Review*, 21(1), 1-13.

[48] Phillips, P. C. B. (2010): "The Mysteries of Trend," Cowles Foundation Discussion Paper No. 1771.

[49] Prebisch, R. (1950): "The Economic Development of Latin America and its Principal Problems," Reprinted in: *Economic Bulletin for Latin America*, 7, 1962, pp. 1-22.

[50] Scheinkman, J. A. and J. Schechtman (1983): "A Simple Competitive Model with Production and Storage," *Review of Economic Studies*, 50, 427-441.

[51] Singer, H. (1950): "The Distribution of Gains between Investing and Borrowing Countries," *American Economic Review*, 40, 473-485.

[52] Stokey, N. L., R. E. Lucas, JR., and E. C. Prescott (1989): *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

[53] Stocks, J. H., J. Wright and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," J*ournal of Business & Economic Statistics*, 20(4), 518-529.

[54] Tauchen, G. (1986): "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business & Economic Statistics*, 4(4), 397-416.

[55] West, K. D. (2002): "Efficient GMM Estimation of Weak AR Processes," *Economic Letters*, 75, 415-418.

[56] West, K. D., K. F. Wong, and S. Anatolyev (2009): "Instrumental Variables Estimation of Heteroskedastic Linear Models using All Lags of Instruments," *Econometric Reviews*, 28(5), 441-467.

[57] Williams, J. C. and B. D. Wright (1991): *Storage and Commodity Markets*, Cambridge: Cambridge University Press.

[58] Wright, B. D. and J. C. Williams (1982): "The Economic Role of Commodity Storage," *Economic Journal*, 92(367), 596-614.

[59] Wright, B. D. and J. C. Williams (1984): "The Welfare Effects of the Introduction of Storage," *The Quarterly Journal of Economics,* 99(1), 169-192.

Appendix A

Mathematical Proof for Chapter 1

Proof. Let $p_{II}(\cdot)$ denote the unique function that satisfies the Euler equation of Model II:

$$p_{II}(x_{II}) = \max\left\{f_{II}(x_{II}), \frac{1}{\gamma}Ep_{II}\left((1-\rho)\left(x_{II} - f_{II}^{-1}(p_{II}(x_{II}) + \omega_{II})\right)\right)\right\}.$$

Let $g(x_{II}) \equiv \lambda p_I(x_I)$. If

$$f_{II}(x_I) = \lambda f_I(\lambda f_I(x_I)).$$

Thus,

$$\begin{aligned} &\max\left\{f_{II}(x_{II}), \frac{1}{\gamma}Ep_{II}\left((1-\rho)\left(x_{II} - f_{II}^{-1}(p_{II}(x_{II}) + \omega_{II})\right)\right)\right\} \\ =&\max\left\{\lambda f_{II}(x_{II}), \frac{1}{\gamma}Eg\circ\Gamma\left((1-\rho)\left(x_I - f_I^{-1}(p_I(x_I) + \omega_I)\right)\right)\right\} \\ =&\max\left\{\lambda f_I(x_I), \frac{\lambda}{\gamma}Ep_I\left((1-\rho)\left(x_I - f_I^{-1}(p_I(x_I) + \omega_I)\right)\right)\right\} \\ =&\lambda p_I(x_I) = g(x_I). \end{aligned}$$

By uniqueness of $p_{II}(\cdot)$,

$$g(x_{II}) = f_{II}(x_I) = \lambda f_I(x_I).$$

Then,

$$\begin{aligned} f_I^{-1}(p_{II}) &= \Gamma f_I^{-1}(p_I), \\ p_{II}^{-1}(p_{II}) &= \Gamma p_I^{-1}(p_I). \end{aligned}$$

Therefore,

$$\begin{aligned} \zeta_{II}(p_{II}) &= Var\left(p_{II}\left((1-\rho)\left(p_{II}^{-1}(p_{II}) - f_{II}^{-1}(p_{II})\right) + \omega_{II}\right)\right) \\ &= Var\left(p_{II}\circ\Gamma\left((1-\rho)\left(p_I^{-1}(p_I) - f_t^{-1}(p_I)\right) + \omega_I\right)\right) \\ &= Var\left(\lambda f_I\left((1-\rho)\left(p_I^{-1}(p_I) - f_t^{-1}(p_I)\right) + \omega_I\right)\right) \\ &= \lambda^2\zeta_I(p_I) \end{aligned}$$

□

Appendix B

Further Discussion on “Weak” Instrument

In this appendix, I show that lag-three price does not suffer from the problem of weak instrument/identification as defined in (Stock, Wright and Yogo, 2002). In GMM, the parameter vector θ is identified by the conditional mean restrictions $E\left[u\left(y_t,\theta_0\right)\right]=0$, where θ_0 is the true value for θ and z_t is a vector of instruments; this in turn implies $E\left[u\left(y_t,\theta_0\right)\otimes z_t\right]=0$. If the instruments are relevant, then $E\left[u\left(y_t,\theta\right)\otimes z_t\right]\neq 0$ for $\theta\neq\theta_0$, a necessary condition for θ to be identified. If $E\left[u\left(y_t,\theta\right)\otimes z_t\right]$ is nearly 0 for $\theta\neq\theta_0$, then θ is weakly identified. The current problem is different. Indeed, even lag-three price does not suffer from weak identification with plentiful data. To show this, I plot in Figure B.1 the large sample approximant to $E\left[u_t\left(\theta\right)p_{t-3}\right]$ in a neighborhood around the true parameter. The surface has significant negative gradient in both directions. Although the gradient in the direction of p^* is less negative than in the direction of γ (as evident from the contour plot beneath the surface), it is far from being nearly zero.

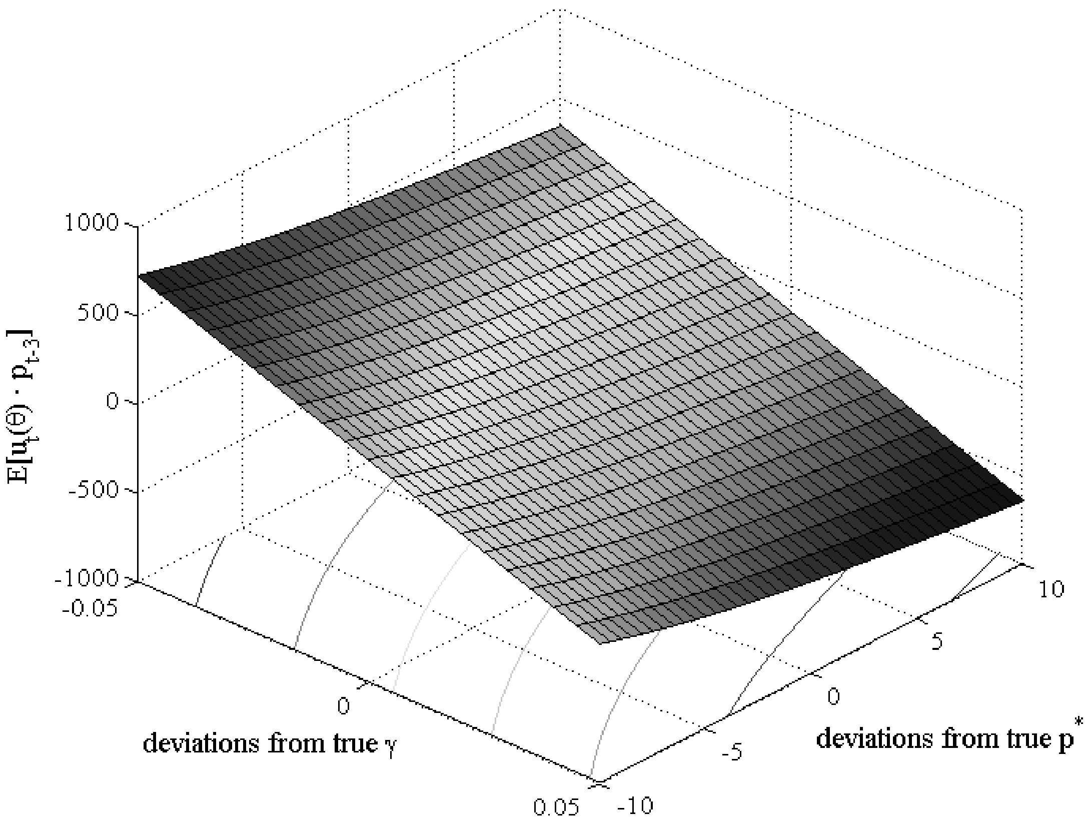

Figure B.1: Large sample approximation to $E\left[u_t\left(\theta\right)p_{t-3}\right]$

Appendix C

Further Discussion on Choosing Preliminary Estimator

Here, I briefly investigate the relation between the performances of the proposed method and the preliminary estimator. In the following table, I present the final performances when various preliminary estimators are used, assuming that the auxiliary model is the same as the true. The preliminary estimator in the last column uses instrument $[1, p_{t-1}]$ and identity weighting matrix. The simulated sample used here is exactly the same as in the main context. For convenience, I also copy down the performance of AOIV0 when BIV is used in the preliminary step.

Regarding estimating γ, the performance when using BIV as the preliminary estimator is obviously superior over the other two under the RMSE criterion. Its 25 and 75 percent values are also closest to the true though its median is bigger than that when DL1 is used.

Regarding estimating p^*, using BIV is better than using DL1 but worse than using $[1, p_{t-1}]$ under the RMSE criterion. It is interesting, however, to observe that from mean and the percentile values, using $[1, p_{t-1}]$ tends to lower estimate p^*, but its STD and RMSE are smaller. This suggests that using BIV or DL1 tends to generate fatter tails. Again, like the comparison between BIV and AOIVi in the main context, if bounds of p^* is imposed in estimation to forbid the estimates going into unrealistic regions, using BIV may outperform using $[1, p_{t-1}]$. Imposing bounds may also improve the performance when using DL1 in the preliminary step, but it is unclear that how much this can reduce its relatively large RMSE.

Table C.1: The performance of the proposed method w.r.t. various preliminary estimators

Preliminary estimator:		BIV	DL1	$[1, p_{t-1}]$
	Mean	1.0537	1.0524	1.0544
	25%	1.0419	1.0381	1.0409
γ	50%	1.0517	1.0503	1.0520
$(= 1.05)$	75%	1.0629	1.0642	1.0649
	STD	0.0164	0.0197	0.0188
	RMSE	0.0168	0.0199	0.0193
	Mean	113.095	113.809	112.245
	25%	107.336	107.688	106.434
p^*	50%	112.558	113.092	111.654
$(= 114.1243)$	75%	118.137	119.064	117.402
	STD	8.395	8.927	8.141
	RMSE	8.485	8.932	8.355

Note. 10000 replications with sample size 100.

Appendix D

Robustness Check of Estimation Result of Chapter 2

Table D.1 to D.6 present estimation results with respect to number of iterations for all specifications and $\mu = 0.85,\ 0.90$. For I and II result becomes stable and robust to μ after 3 iterations. The general pattern is the same for III, except that the convergence is slower when $\mu = 0.9$. For any given number of iteration and μ, estimate for γ is smaller under I than under II and bigger than under III, suggesting that the order of the estimated γ is a result of the systematic difference among the three estimators.

Table D.1: Sensitivity to number of iterations, specification I, $\mu = 0.85$

# Iteration:	γ	ζ^*	λ	α	S.O.	J_T
0	1.0037	1.9375	0.9757	0.4630	7	2.8719
	(0.0267)	(0.1340)	(0.0016)	(0.0476)		(0.9099)
1	1.0165	1.9290	0.9752	0.4726	7	2.5319
	(0.0253)	(0.1329)	(0.0016)	(0.0479)		(0.8884)
2	1.0183	1.9254	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)
3	1.0186	1.9249	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)
4	1.0186	1.9248	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)
5	1.0186	1.9248	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)

Note. Standard errors in parentheses. In the final column, probability values in parentheses. J_T stands for the over-identifying test value.

Table D.2: Sensitivity to number of iterations, specification I, $\mu = 0.90$

# Iteration:	γ	ζ^*	λ	α	S.O.	J_T
0	1.0059	1.9204	0.9756	0.4632	7	2.8181
	(0.0268)	(0.1338)	(0.0016)	(0.0477)		(0.9068)
1	1.0168	1.9284	0.9752	0.4726	7	2.5307
	(0.0253)	(0.1328)	(0.0016)	(0.0479)		(0.8883)
2	1.0184	1.9253	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)
3	1.0186	1.9249	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)
4	1.0186	1.9248	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)
5	1.0186	1.9248	0.9753	0.4725	7	2.5277
	(0.0253)	(0.1326)	(0.0016)	(0.0479)		(0.8881)

Note. Standard errors in parentheses. In the final column, probability values in parentheses. J_T stands for the over-identifying test value.

Table D.3: Sensitivity to number of iterations, specification II, $\mu = 0.85$

# Iteration:	γ	ζ^*	λ	α	S.O.	J_T
0	1.0286	1.9368	0.9751	0.4763	7	2.3199
	(0.0155)	(0.1063)	(0.0022)	(0.0565)		(0.8723)
1	1.0382	1.9368	0.9751	0.4763	7	2.1557
	(0.0146)	(0.1047)	(0.0022)	(0.0565)		(0.8580)
2	1.0393	1.9368	0.9751	0.4763	7	2.1515
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8576)
3	1.0394	1.9368	0.9751	0.4763	7	2.1512
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8575)
4	1.0394	1.9368	0.9751	0.4763	7	2.1511
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8575)
5	1.0394	1.9368	0.9751	0.4763	7	2.1511
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8575)

Note. Standard errors in parentheses. In the final column, probability values in parentheses. J_T stands for the over-identifying test value.

Table D.4: Sensitivity to number of iterations, specification II, $\mu = 0.90$

# Iteration:	γ	ζ^*	λ	α	S.O.	J_T
0	1.0186	1.9614	0.9751	0.4763	7	2.5073
	(0.0153)	(0.1057)	(0.0022)	(0.0565)		(0.8867)
1	1.0374	1.9368	0.9751	0.4763	7	2.1608
	(0.0146)	(0.1048)	(0.0022)	(0.0565)		(0.8584)
2	1.0392	1.9368	0.9751	0.4763	7	2.1517
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8576)
3	1.0394	1.9368	0.9751	0.4763	7	2.1512
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8575)
4	1.0394	1.9368	0.9751	0.4763	7	2.1511
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8575)
5	1.0394	1.9368	0.9751	0.4763	7	2.1511
	(0.0146)	(0.1045)	(0.0022)	(0.0565)		(0.8575)

Note. Standard errors in parentheses. In the final column, probability values in parentheses. J_T stands for the over-identifying test value.

Table D.5: Sensitivity to number of iterations, specification III, $\mu = 0.85$

# Iteration:	γ	ζ^*	λ	α	S.O.	J_T
0	1.0030	1.3756			7	0.2103
	(0.0297)	(0.0883)				(0.3535)
1	1.0084	1.3684			7	0.1733
	(0.0268)	(0.0861)				(0.3228)
2	1.0084	1.3683			7	0.1733
	(0.0268)	(0.0861)				(0.3228)
3	1.0084	1.3683			7	0.1733
	(0.0268)	(0.0861)				(0.3228)
4	1.0084	1.3683			7	0.1733
	(0.0268)	(0.0861)				(0.3228)
5	1.0084	1.3683			7	0.1733
	(0.0268)	(0.0861)				(0.3228)

Note. Standard errors in parentheses. In the final column, probability values in parentheses. J_T stands for the over-identifying test value.

Table D.6: Sensitivity to number of iterations, specification III, $\mu = 0.90$

# Iteration:	γ	ζ^*	λ	α	S.O.	J_T
0	1.0000	1.4477			3	2.0265
	(0.0275)	(0.1221)				(0.8454)
1	1.0000	1.5814			2	1.7716
	(0.0245)	(0.1259)				(0.8168)
2	1.0000	1.6389			1	1.8949
	(0.0240)	(0.1856)				(0.8313)
3	1.0075	1.3669			7	0.1752
	(0.0268)	(0.0861)				(0.3245)
4	1.0084	1.3683			7	0.1733
	(0.0268)	(0.0861)				(0.3228)
5	1.0084	1.3683			7	0.1733
	(0.0268)	(0.0861)				(0.3228)

Note. Standard errors in parentheses. In the final column, probability values in parentheses. J_T stands for the over-identifying test value.