# Bayesian Learning, Shutdown and Convergence *

Leopold Sögner †

October 15, 2011

**Abstract**

This article investigates a partial equilibrium production model with dynamic information aggregation. Firms apply Bayesian learning to estimate the unknown model parameter. In the baseline setting, where prices and quantities are supported by the real line and the noise term is Gaussian, convergence of the limited information to the full information setting is obtained. Imposing a non-negativity constraint on quantities destroys the convergence results obtained in the baseline model. With the constraints firms learning an unknown demand intercept parameter exit with strictly positive probability, even when the true value of this parameter would induce production in the full information setting. Parts of the model can be rescued by assuming bounded support for the stochastic noise term. Although, shut-down can be excluded with bounded support when the forecasts of the agents satisfy relatively mild regularity conditions, Bayesian learning and convergence need not take place in general.

**Keywords:** Bayesian Learning, Consistency, Convergence.

**JEL:** D82, G10.

*The author appreciates helpful comments from Larry Blume, Egbert Dierker, Hildegard Dierker, Klaus Ritzberger, Andreas Danis, David Florysiak, Michael Greinecker, Maarten Janssen, Julian Kolm, Martin Meier and all participants of the Micro Jour Fixe at IHS (summer & winter term 2010), the VGSE seminar (Vienna 2011) and the $11^{th}$ SAET conference (Faro 2011).

†Leopold Sögner, Department of Economics and Finance, Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna, Austria, soegner@ihs.ac.at

# 1 Introduction

One of the workhorse models when considering learning is a partial equilibrium setting with normally distributed noise and agents applying Bayesian learning. The mean parameter of the stochastic noise term is unknown while the variance term is kept fixed. By assuming a conjugate normal prior, the posterior of the unknown mean parameter is normal. The posterior mean becomes a mixture of the prior mean and the sample average. In addition to some other nice features of the normal distribution, this property also motivates why a normal setting is so frequently used: A dynamic setting with affine implied law of motion results in an affine perceived actual law of motion.

In this article the baseline model is a partial equilibrium model with production as investigated in Jun and Vives (1996); *JV* in the following. *JV* have already shown that learning is not equivalent to convergence of the limited information to the full information equilibrium. However, for all non-unit root settings the unknown parameter is learned, while convergence of the limited information quantities to full information quantities is attained for all feasible parameter values.

This paper shows that different deviations from the affine normal setting need neither result in learning nor in convergence. While quantities are real numbers in the baseline model, by requiring non-negative quantities – while sticking to a normally distributed demand intercept – we show that due to a shut-down option limited informed firms could leave the market while informed firms stay in there. Here the convergence result of the baseline setting breaks down.

Let us relate this article to recent literature: As already stated above *JV* have already shown that learning and convergence are not equivalent. The *JV* setting assumes a continuum of price takers and public signals only. By the first assumption, solving for equilibrium does not require concepts from game theory, while the second assumption excludes any herding effects (for an overview see e.g. Vives (2008)[chapter 6]). Vives (2008) also presents a convergence result with herding behavior; also this model works with the normal distribution. Using the deviations discussed in this paper will also have impacts on the convergence results presented there. Alternatively, Smith and Sørensen (2000) considered a herding model with binomial outcomes. If there were no private signals in their setting, the Bayes estimator would be consistent since the sample space is finite. Smith and Sørensen (2000) show that due to learning from public and private information *confounded learning* can appear, which is a situation where a Markov transition probability becomes independent of the last realization of the state variable. By the fact that only public information arises in our model, such a situation cannot appear in our model. However, our setting provides different examples where learning does not take place. We neither need heterogeneity of agents (informed vs. non-informed) nor herding behavior. Non-convergent behavior either arises from deviations from the normal distribution or from a shut-down condition (or both). For further stability results with respect to herding we refer the reader to Alfarano and Milakovic (2009).

A model very similar to our baseline setting has also been investigated in Guesnerie (1992) to check for *eductive stability* or a *strongly rational expectations equilibrium*; see also Guesnerie (1993) and Guesnerie and Jara-Moroni (2009) for a more general treatment of this concept. This strand of literature raises the question whether rational expectations can be supported by Game theoretic reasoning. Here the author has shown that strong rationality depends on the eigenvalue of a cobweb function derived from a composition of the inverse aggregate demand with the supply function. In this article we abstract from the problem of strong rationalization and put the focus on the requirement of non-negative quantities.

This paper is organized as follows: Section 2 describes the benchmark model. Section 3 introduces non-negative quantities and a shutdown option for firms. If the feasible set is restricted to the non-negative real axis, shut-down - that is to say a production of zero - can be an optimal strategy. Since no signals are received after shut-down, this implies that the agents also decide to stop learning. Then even in the limit the unknown parameter remains a random variable. With these modifications the convergence results obtained by $JV$ may break down.

Section 4 shows that a bounded support on the stochastic noise term *and* regularity conditions on the forecasts provide a more realistic model in which the results of the benchmark model can be resurrected. Although a convergence result will follow easily in a quite mechanical way, it is important to note that this simple setup already shows that even with bounded support of the noise term, convergence need not be attained in general. Regularity conditions on the learning scheme are necessary to guarantee convergence. Examples show that matching these conditions can but need not be trivial. Last but not least we investigate robustness by allowing for deviations from normal noise. We present examples - based on Diaconis and Freedman (1986) and Christensen (2009) - where convergence need not occur even with bounded support. This last topic is related to the consistency of the Bayes estimate (see e.g. Diaconis and Freedman (1986) or Strasser (1985)). Applications and extension in economic theory have been presented e.g. in Blume and Easley (1993), Blume and Easley (1998), Feldman (1991), Kalai and Lehrer (1992), Kalai and Lehrer (1992) and Stinchcombe (2005). Appendix B will reconsider the results presented in literature and provides some further examples.

## 2 The Benchmark Model

This section describes the model and the key results obtained in *JV*. Their results are based on the assumption that both prices, $p_t$, and the quantities, $x_t$, in particular may be negative. *JV* and also Vives (2008)[chapter 7.2] consider a discrete time model of dynamic information aggregation; time is indexed by $t = 0, 1, 2, \ldots$, economic activity starts at $t = 1$. The aggregate *demand* function is described by the stochastic linear relationship:

$$p_t = z_t - \beta x_t \text{ for } t \geq 1 \ . \tag{1}$$

$p_t$ is the price established in period $t$, $\beta > 0$ is a constant and $x_t$ is the aggregate quantity consumed. $z_t$ is a random variable. The *fractional differences* $\Delta^\zeta z_t = z_t - \zeta z_{t-1}$ are described by

$$\Delta^\zeta z_t = (1-\zeta)\theta + \eta_t \ , \tag{2}$$

where $\theta, \zeta \in \mathrm{R}$. With $u_t = z_t - \theta$ we get an autoregressive and a moving average representation:

$$\begin{aligned} z_t &= \zeta z_t + (1-\zeta)\theta + \eta_t \ , \\ &= \theta + u_t = \theta + \sum_{s=0}^{t-1} \zeta^s \eta_{t-s} + u_0 = \theta + \sum_{s=0}^{t-1} \zeta^s \eta_{t-s} + (z_0 - \theta) \ . \end{aligned} \tag{3}$$

In addition $u_t = \zeta u_{t-1} + \eta_t$. In the baseline setting $\eta_t$ is *iid* normal with mean zero and variance $\sigma^2$, $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$. For $\zeta \neq 1$, the process follows a first order autoregressive process with normal innovations, while for $\zeta = 1$ we get a random walk. By this specification $x_t$ and $p_t \in \mathrm{R}$. $z_t$ is the stochastic demand intercept.[1] We assume that the process $(z_t)$ is started at $z_0 = \theta$, which has also be done implicitly in *JV* by assuming $u_0 = 0$. This assumption will reduce the computational burden in Section 4. It is important to note that $z_0$ cannot be observed by the firms.

Consider a continuum of *firms* $i \in [0, 1]$ endowed with Lebesque measure. Each firm $i$ is a price taker and produces the homogeneous output $x_{it}$ at cost $C(x_{it}) = \frac{\lambda}{2} x_{it}^2$; the parameter $\lambda > 0$. Aggregate output fulfills $x_t = \int_0^1 x_{it} di$.[2] Depending on what firms observe, distinguish between:

1. *Full information (FI)*: At period $t$, the firms know $\theta$ and the past prices $p^{t-1} = \{p_1, \ldots, p_{t-1}\}$. In addition firms know the structure of demand (1) and the model parameters $\lambda, \beta, \sigma_\eta^2, \zeta$, and $u_0 = 0$.

2. *Limited information (LI)*: Firms know $\lambda, \beta, \sigma_\eta^2, \zeta$, and $u_0 = 0$ and observe past prices $p^{t-1}$. They *do not* know $\theta$, but the structure of demand (1) and (2). Firms use Bayes' rule to update beliefs about $\theta$. The prior is given by $\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$. This is what Vives (2008)[chapter 7.1] calls *learning within an equilibrium.*

For notational simplicity abbreviate the information sets by $\mathcal{I}_t^{FI}$ and $\mathcal{I}_t^{LI}$, for the full information and the limited information case, respectively. Since firms are price takers, the profit function is $\pi_{it} = p_t x_{it} - \frac{\lambda}{2} x_{it}^2$ and the value function is $V_{it} = (1-\delta)\mathbb{E}\left(\sum_{k=0}^{\infty} \delta^k \left(p_{t+k} x_{i,t+k} - \frac{\lambda}{2} x_{i,t+k}^2\right) | \mathcal{I}_t^{(.)}\right)$ with some discount factor $\delta \in (0, 1)$. By (1), (3) and $\zeta \in \mathrm{R}$ the value function need not be finite in general. However, by the structure of the optimization problem maximizing $V_{it}$ breaks up into the one period optimization problems

$$\max_{x_{it}} \quad \mathbb{E}\left(p_t x_{it} - \frac{\lambda}{2} x_{it}^2 | \mathcal{I}_t^{(.)}\right) . \tag{4}$$

Given the information set $\mathcal{I}_t^{(.)}$ the first order condition yields:

$$x_{it} = \frac{\mathbb{E}\left(p_t | \mathcal{I}_t^{(.)}\right)}{\lambda} . \tag{5}$$

[1]For some examples it is more convenient the work with $\Delta^\zeta z_t = \varepsilon_t$, where $\varepsilon_t$ is centered around $(1-\zeta)\theta$ (if the mean exists). Such a transformation makes sense if asymmetric noise (e.g. asymmetric (truncated) normal) or only noise on a subset of R will be considered (e.g. Gamma distribution, truncated distributions). This notation will be applied in Examples 1 to 4 and in some examples in Appendix B.

[2]Since $x_{it}$ are equal for all $i$, except for a countably many $i$ - the exact law of large numbers can be applied, such that $x_t = \int_0^1 x_{it} di$ still holds. For more details see Sun (2006).

*Market Clearing*: In period $t$ firm $i$ produces $x_{it} = \frac{\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)}{\lambda}$ and aggregate (average) output is $\int_0^1 x_{it}di = x_t = \frac{\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)}{\lambda}$. Immediately after this output decision, the random variable $\eta_t$ realizes, resulting in $p_t = z_t - \beta x_t = z_t - \beta\frac{\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)}{\lambda}$.

**Remark 1.** Since $sgn(x_{it}) = sgn(\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right))$ by (5), expected revenues are positive no matter whether positive or negative prices are expected. Substitution of (5) into the expected profit function results in

$$\mathbb{E}\left(\pi_{it}|\mathcal{I}_t^{(\cdot)}\right) = \frac{\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)^2}{\lambda} - \frac{\lambda\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)^2}{2\lambda^2} = \frac{\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)^2}{2\lambda} > 0 \quad (a.s.)\ .$$

Agents commit to supply $x_{it} \in \mathrm{R}$ even if $sgn(x_{it}) = sgn(\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)) \neq sgn(p_t)$. In this case $\pi_{it} < 0$. Hence, realized profits $\pi_{it} \in \mathrm{R}$. By the law of iterated expectations $\mathbb{E}\left(\pi_{it}\right) > 0$.

*Convergence, Prediction and Learning*: We restrict our analysis to *learning within an equilibrium*, as defined in Vives (2008)[p. 249]. Examples for different learning schemes in different fields of economics are provided e.g. in Timmermann (1996), Brock and Hommes (1997), Kelly and Kolstad (1999), Routledge (1999) or Evans and Honkapohja (2001). For an overview and a bulk of literature we refer the reader to Vives (2008)[chapters 7.1 and 10.2].

Due to (5), firms have to predict the price $p_t$. By the inverse demand function (1) we get

$$\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right) = \mathbb{E}\left(z_t|\mathcal{I}_t^{(\cdot)}\right) - \beta x_t\ . \tag{6}$$

$x_t$ is $\mathcal{I}_t^{(\cdot)}$ measurable. For $k \geq 0$, the $k$ step ahead prediction is obtained by means of

$$\mathbb{E}\left(z_{t+k}|\mathcal{I}_t^{(\cdot)}\right) \quad = \quad (1-\zeta)\mathbb{E}\left(\theta|\mathcal{I}_t^{(\cdot)}\right)\sum_{i=0}^{k}\zeta^i + \zeta^{k+1}z_{t-1}\ . \tag{7}$$

For the full information case, where $\theta$ is known, (7) results in $\mathbb{E}\left(z_t|\mathcal{I}_t^{FI}\right) = (1-\zeta)\theta + \zeta z_{t-1}$, such that

$$x_t^{FI} = \frac{(1-\zeta)\theta + \zeta z_{t-1}}{\lambda+\beta} = \frac{\theta + \zeta u_{t-1}}{\lambda+\beta}\ . \tag{8}$$

By the inverse demand function $y_t := p_t + \beta x_t = p_t + \beta\frac{\mathbb{E}\left(p_t|\mathcal{I}_t^{(\cdot)}\right)}{\lambda} = \theta + u_t = z_t$. With limited information firms know that $y_t = z_t$. From (1) and the specification of the noise term, $\eta_t \sim\ iid\ \mathcal{N}(0, \sigma_\eta^2)$, algebra yields

$$z_t = \theta + \zeta(z_{t-1} - \theta) + \eta_t \text{ and } \Delta^\zeta z_t := z_t - \zeta z_{t-1} = (1-\zeta)\theta + \eta_t \ . \tag{9}$$

For $\zeta = 1$, we directly observe from $z_t = \theta(1-\zeta) + \zeta z_{t-1} + \eta_t$ that the parameter $\theta$ is not identified. That is to say, for any $z_t \in \mathrm{R}$ the likelihood $f(z_t|\theta, \zeta = 1) = f(z_t|\theta', \zeta = 1)$ for arbitrary pairs $\theta, \theta'$.

The last term in (9) corresponds to a linear regression setting with response variable $\Delta^\zeta z_t$, the (constant) prediction variable $(1-\zeta)$ and normal innovations $\eta_t$. With the conjugate normal prior[3] $\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$, under limited information firms derive the posterior distribution of the parameter $\theta$ by means of (see e.g. Vives (2008)[Appendix 10.2], Chib (1993), Robert (1994)[Chapter 4]):

$$\theta \ \sim \ \mathcal{N}(a_t, A_t) \text{ where } a_t = \frac{(\frac{1}{\sigma_\eta^2}(1-\zeta)\sum_{s=1}^{t} \Delta^\zeta z_s + \frac{\theta_0}{\sigma_0^2}}{(\frac{1}{\sigma_\eta^2}(1-\zeta)^2 t + \frac{1}{\sigma_0^2}} \ , A_t = \left(\frac{1}{\sigma_\eta^2}(1-\zeta)^2 t + \frac{1}{\sigma_0^2}\right)^{-1} . \tag{10}$$

By means of $a_t$ firms immediately get the conditional expectation of $\theta$.[4] Since $\mathcal{I}_t^{LI}$ is generated by observations from $1, \ldots, t-1$, the conditional expectation $\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right) = a_{t-1}$. Then, by (7) forecasts of $z_{t+k}$ are

$$\mathbb{E}\left(z_{t+k}|\mathcal{I}_t^{LI}\right) = (1-\zeta)\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right)\sum_{i=0}^{k}\zeta^i + \zeta^{k+1} z_{t-1} = (1-\zeta)a_{t-1}\sum_{i=0}^{k}\zeta^i + \zeta^{k+1} z_{t-1} \ . \tag{11}$$

(5) and (11), with $k = 0$, yield

$$x_t = \frac{\mathbb{E}\left(z_t|\mathcal{I}_t^{LI}\right)}{\lambda + \beta} \ . \tag{12}$$

Before we restate the *JV* convergence result for the baseline model, let us briefly discuss some properties of $(x_t)$ and $(x_t^{FI})$. $(x_t^{FI})$ follows a first order autoregressive process. Generally, a first order autoregressive process $u_t = \zeta u_{t-1} + \eta_t$ is ergodic (stationary in the limit) if $|\zeta| < 1$ and $\mathbb{E}(|\eta_t|) < \infty$ (see e.g. Meyn and

[3]The following analysis can also be performed with an uninformative Jeffreys prior (see e.g. Robert (1994)), which can be motivated by the (frequentist) argument that information is only providing by the data. Minimizing the impact of the prior is done with this type of prior. For the current normal setting this implies that terms including $\theta_0$, $\sigma_0^2$ or both vanish in (10). Proposition 1 and the results in Examples 5 and 6 still hold when replacing the conjugate normal prior by an uninformative Jeffreys prior.

[4]Only $z_t$ for $t \geq 1$ are available. At $t = 1$ firms use $\mathbb{E}\left(z_1|\mathcal{I}_1^{LI}\right) = \theta_0$. At $t = 2$, $z_1$ is already known by the firms, but they cannot calculate $\Delta^\zeta z_1$. Here we assume that firms apply $\Delta^\zeta z_1 = z_1 - \theta_0$. Then (10) can be used. Alternatively, we could also start with a setup where no $\Delta^\zeta z_1$ is available but agents apply Bayesian simulations methods to sample the joint posterior of the parameter $\theta$ and the initial value as described in Appendix C.

Tweedie (2009)[Chapters 11 and 15]; in this case $(\eta_t)$ is also geometrically ergodic). This implies that the distribution of $u_t$, $\mu_t$, converges to its stationary distribution $\mu$. If $u_0 \sim \mu$, then the process is stationary and ergodic. $u_t \in L^p$ if $\int |u_t|^p d\mu_t < \infty$; where $\mu_t$ is the probability law of $u_t$. For normal innovations, if $|\zeta| < 1$ then $(u_t)$ is stationary and $p$-th movements exist such that $u_t \in L^p$. Therefore, $(x_t^{FI})$ is stationary in the limit and in $L^p$ if $|\zeta| < 1$. If $|\zeta| \geq 1$, neither $(x_t)$ nor $(x_t^{FI})$ are ergodic. Discussing $L^p$ convergence with quantities does not make sense in this case. If $|\zeta| \geq 1$, we can work with the *fractional differences* $\Delta^\zeta x_t^{(.)} = x_t^{(.)} - \zeta x_{t-1}^{(.)}$. For the full-information quantities this results in $\Delta^\zeta x_t^{FI} = \frac{(1-\zeta)[\theta+\zeta\eta_{t-1}]}{\lambda+\beta}$. $(\Delta^\zeta x_t^{FI})$ is stationary/$\in L^p$ if $(\eta_t)$ is stationary/$\in L^p$. Given our model assumptions, the fractional differences $\Delta^\zeta x_t^{FI}$ are ergodic and $\in L^p$ for all $\zeta$. For the limited information case, some algebra yields

$$\Delta^\zeta x_t^{LI} = \Delta^\zeta x_t^{FI} + \frac{(1-\zeta)[a_{t-1} - \zeta a_{t-2} - (1-\zeta)\theta]}{\lambda+\beta} \ . \tag{13}$$

The third term in (13) goes to zero if $(a_t)$ converges to $\theta$. (13) also implies that $x_t - x_t^{FI} = \frac{1-\zeta}{\lambda+\beta}(a_{t-1} - \theta)$. If $\zeta = 1$, then $x_t^{LI} = x_t^{FI}$ at least for $t \geq 2$, but $\theta$ is not learned. For all $\zeta \neq 1$, $(a_t)$ converges to $\theta$ almost surely and in $L^p$. $\Delta^\zeta x_t^{LI}$ and $x_t^{LI}$ converge to their full information counterparts, for $\zeta \in \mathrm{R}$ and $|\zeta| < 1$, respectively. Based on this discussion we reformulate the convergence result obtained by *JV*:

**Proposition 1.** *[Propositions 2.1, 2.2 in* JV*; Proposition 7.1 in Vives (2008)]*

*(i) If $|\zeta| < 1$ then $\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right)$ converges to $\theta$ and $x_t \to x_t^{FI}$ (a.s. and in mean square).*

*(ii) If $\zeta = 1$, except for the first period, no information about $\theta$ can be inferred from prices (with the precision of $\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right)$ constant at $1/\sigma_0^2 + 1/\sigma_\eta^2$) but $x_t = x_t^{FI}$ for $t \geq 2$.*

*(iii) If $\zeta \in \mathrm{R}$ and $\zeta \neq 1$, then $\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right)$ converges to $\theta$, $\Delta^\zeta x_t \to \Delta^\zeta x_t^{FI}$ (a.s. and in mean square).*

**Remark 2.** *JV* also derived that $\sqrt{t}$ convergence is attained. This issue will not be investigated here. If $\zeta \neq 1$ then $\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right) \to \theta$. This implies $\Delta^\zeta x_t \to \Delta^\zeta x_t^{FI}$. Due to the second part of Proposition 1, this is only an "if" statement and not "if and only if". That is to say convergence of $x_t$ and learning $\theta$ are not equivalent. Subfigures (a) and (b) of Figure 1 will provide a graphical illustration of the convergence result obtained in this section.

**Remark 3.** When reconsidering the model and the results of this section, $\theta$ is estimated to perform predictions of prices and $z_t$. One might therefore claim that - at least from the firms' point of view - the ultimate goal is not learning but forecasting. The Bayesian tool to describe the distribution of the forecast is the predictive density (predictive distribution). It is derived by means of the Bayes theorem (see e.g. Bickel and Doksum (2001)[p. 254] and Poirier (1995)[Chapter 8.6]).

Generally estimation and prediction are not equivalent. Blume and Easley (1998) present an example where learning occurs but the predictive density remains the same while learning occurs, and an example where the predictive density remains the same even if learning does not take place. In addition the authors show that "the marginal distribution of the parameter absolutely continuous with respect to the likelihood" is sufficient for convergence of the predictive distribution (see Blume and Easley (1998)[Theorem 2.3]).

For the current model: If $\zeta \neq 1$ the predictive density of $z_t - \zeta z_{t-1}$ is a normal distribution with mean $(1-\zeta)a_{t-1}$ and variance $(1-\zeta)^2 A_{t-1} + \sigma_\eta^2$. If $t \to \infty$ then $a_t \to \theta$ and the forecast variance converges to $\sigma_\eta^2$. This implies that the asymptotic Bayesian and frequentist prediction intervals are equal. (If the parameter $\sigma_\eta^2$ is unknown, the predictive density is a t-distribution. The convergence result with the forecast densities still holds.) For $\zeta = 1$ the predictive distribution remains the same for all $t \geq 2$ although $\theta$ is not learned. $z_t$ is normal with mean $z_{t-1}$ and variance $\sigma_\eta^2$. That is to say, although learning need not take place in this baseline model, the forecasting distribution converges to the correct limit distribution for all $\zeta$.

*Robustness of the Linear Gaussian Model*: A further question arises when we deviate from the Gaussian distribution. In general not every Bayes estimator has to be consistent, if we change the distribution of $\eta_t$. This implies that $\theta_t$ need not converge to $\theta$ in an arbitrary setting even if $\zeta = 0$. Appendix B presents some statistical theory and the convergence results from literature. Sufficient conditions for convergence in regular and non-regular cases are provided (i.e. where the support of the distribution is part of the unknown parameter vector). Dirichlet mixtures provide examples to break down the convergence results obtained with the linear Gaussian setting presented in this section. In these examples the Bayes estimate is not consistent on sets of prior measure zero (small in a measure theoretic sense), while the same set is of Baire category one (not small in a topological sense); for more details see Examples 5 and 6 in

Appendix B.

## 3 Shutdown and Non-negative Quantities

Section 2 assumed that the domains of $x_t$ and $p_t$ are the real line. What does this mean in economic terms? E.g. $x_t = x_{it} < 0$ implies that firms consume the product/ use $x_t$ as an input and consumers supply $x_t$. If $x_t = x_{it} \neq 0$ a cost of $\frac{\lambda}{2}x_{it}^2$ arises for all firms, no matter whether firms produce or consume. In addition, if prices and quantities are negative this implies that the firm receives a subsidy for producing a negative quantity. Therefore, the assumptions of Section 2 are not very convincing whenever quantities and prices are outside the non-negative orthant. One may argue that the model parameters can be chosen such that the probability of negative quantities and prices becomes small. Indeed this might be the case, but with normally distributed noise there is always a non-zero probability to violate non-negativity constraints.

Now assume $x_{it} \geq 0$. Otherwise stay as close as possible to the model by *JV*. In particular, (1) is maintained.[5] This scenario is motivated as follows: If firms produce $x_{it} = x_t > 0$, but $u_t$ will be realized such that there is no demand for this quantity (negative demand intercept), then firms cannot store this quantity and have negative revenues $p_t x_{it} < 0$, e.g. due to a scrapping cost for destroying $x_{it}$. Formally, the following assumptions are adopted:

**Assumption 1.** Quantities $x_{it} \geq 0$.

**Assumption 2.** Firms are permitted to set $x_{it} = 0$ (*shutdown* in period $t$). If $x_t = x_{it} = 0$, firms do not receive any signal from the market. No price is realized if there is no market.

By Assumption 1, $x_t = \int_0^1 x_{it} di \geq 0$. The domain of $x_t$ is the non-negative part of the real axis. If $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ then $p_t \in \mathrm{R}$ by (1). Assumption 2 could also be included in Section 2, since the expected profits are non-negative there ($x_t = 0$ with a probability of zero), so that shutdown was not an issue. In this section the analysis is more convenient with

**Assumption 3.** $|\zeta| < 1$.

[5] If (1) would be replaced by $p_t = \max(0, \theta + u_t - \beta x_t)$, then the conditional expectation of $p_t$ is no more linear in $x_t$ which makes the model much more complicated.

Assumption 3 excludes the unit root case and explosive processes. Now we stick to Assumptions 1 to 3. This yields:[6]

1. If $x_t > 0$, then $p_t = z_t - \beta x_t$ and $y_t = z_t$ (this last equality holds for all $x_t$ in Section 2).

2. If $x_t = 0$ - by Assumption 2 - firms do not receive any information on the demand intercept $z_t$. In terms of econometrics this implies that $p_t$ and $z_t$ are missing values.

Now $(p_t)$ and $(z_t)$ are time series with some missing observations. Whenever $x_t > 0$, $p_t - \beta x_t = z_t$ continuous to hold. Taking account to missing values, the information sets are $\mathcal{F}_t^{FI}$ and $\mathcal{F}_t^{LI}$.

The conditional expectations are $\mathbb{E}\left(p_{t+k}|\mathcal{F}_t^{(.)}, x_{i,t+k} > 0\right) = \left(\mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{(.)}\right) - \beta x_{i,t+k}\right)\mathbf{1}_{x_{i,t+k}>0}$, the forecasts of $z_{t+k}$ are derived by means of (11), where $\mathcal{I}_t^{(.)}$ has to be substituted by $\mathcal{F}_t^{(.)}$. Regarding parameter estimation the difference to Section 2 is the presence of missing values. Exact Bayesian parameter estimation with missing values is briefly described in Appendix C. In a slightly simpler way we proceed as follows: If at time $t$, $z_{t-j}, \ldots, z_{t-1}$ are missing, then firms adapt (7) to

$$\mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{LI}\right) = (1-\zeta)\mathbb{E}\left(\theta|\mathcal{F}_{t-j}^{LI}\right)\sum_{i=0}^{(k+j)}\zeta^i + \zeta^{k+j+1}z_{t-1} \text{ and } \mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{FI}\right) = \theta + \zeta^{k+j+1}u_{t-j-1} \,. \quad (14)$$

Equipped with (14) firms can solve the profit maximization problem

$$\max_{x_{i,t+k}} \ \mathbb{E}\left(p_{t+k}|\mathcal{F}_t^{(.)}\right)x_{i,t+k} - \frac{\lambda}{2}x_{i,t+k}^2 \ \ s.t. \ x_{i,t+k} \geq 0 \ .$$

The first order conditions are

$$\mathbb{E}\left(p_{t+k}|\mathcal{F}_t^{(.)}\right) - \lambda x_{i,t+k} \leq 0 \text{ and } \left(\mathbb{E}\left(p_{t+k}|\mathcal{F}_t^{(.)}\right) - \lambda x_{i,t+k}\right)x_{i,t+k} = 0 \ . \quad (15)$$

Condition (15) implies that firm $i$ currently produces $x_{it} = \frac{\mathbb{E}\left(p_t|\mathcal{F}_t^{(.)}\right)}{\lambda}$ units if $\mathbb{E}\left(p_t|\mathcal{F}_t^{(.)}\right) > 0$ and 0 otherwise. By (15) firms also obtain the production plans for the periods $t+k$, $k > 0$, given current information. $\mathbb{E}\left(p_{t+k}|\mathcal{F}_t^{(.)}\right) > 0$ requires $\mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{(.)}\right) > 0$ by the model structure (see equations (6) and (11)); if the latter term is positive there is an interior solution. This can be summarized as follows:

[6] From Section 2 we already know that $y_t$ has been defined as $y_t = p_t + \beta x_t$. Whenever production takes place $y_t = p_t + \beta x_t = \theta + u_t = z_t$ has to hold. In Section 2, $y_t = z_t$ for all $t$.

**Proposition 2.** *Suppose that Assumptions 1-3 hold:*

*C1 If* $\mathbb{E}\left(\theta|\mathcal{F}_t^{(.)}\right) > 0$ *and* $\mathbb{E}\left(z_t|\mathcal{F}_t^{(.)}\right) > 0$*. Then in period* $t$*,* $\mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{(.)}\right) > 0$ *for all* $k \geq 0$*. In* $t$ *firms produce* $x_t > 0$*. For periods* $t+k$ *firms currently plan to supply* $x_{t+k} > 0$*.*

*C2 If* $\mathbb{E}\left(\theta|\mathcal{F}_t^{(.)}\right) > 0$ *and* $\mathbb{E}\left(z_t|\mathcal{F}_t^{(.)}\right) < 0$*, then* $\mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{(.)}\right) > 0$ *for some* $k \geq 1$*. Firms do not produce in period* $t$ *but enter the market at period* $t+k$*.*

*C3 If* $\mathbb{E}\left(\theta|\mathcal{F}_t^{(.)}\right) < 0$ *and* $\mathbb{E}\left(z_t|\mathcal{F}_t^{(.)}\right) > 0$*, then firms produce in this period* $x_t > 0$*. At* $t$ *firms plan to exit after* $k > 0$ *periods.*

*C4 If* $\mathbb{E}\left(\theta|\mathcal{F}_t^{(.)}\right) < 0$ *and* $\mathbb{E}\left(z_t|\mathcal{F}_t^{(.)}\right) < 0$*, this also implies* $\mathbb{E}\left(z_{t+k}|\mathcal{F}_t^{(.)}\right) < 0$ *for all* $k \geq 0$*. Firms do not produce and exit for all* $t+k$*,* $k \geq 0$*.*

Proposition 2 has important economic implications: If $\theta > 0$, in the full-information case only production or entry after $k$ periods is possible (cases C1 and C2). However, with limited information cases C3 and C4 are possible as well. In case C3, maybe the firms receive a positive signal to remain in the market. On the other hand, with C4 the following result obtains:

**Corollary 1.** Suppose that $\theta > 0$ and $\mathbb{E}\left(\theta|\mathcal{F}_t^{LI}\right) < 0$ and $\mathbb{E}\left(z_t|\mathcal{F}_t^{LI}\right) < 0$. Then firms with limited information exit and $x_t$ does not converge to $x_t^{FI}$. With $t$ finite, there always exists a $\eta_{t-1}^{+}$ such that $\mathbb{E}\left(\theta|\mathcal{F}_t^{LI}\right) < 0$ and $\mathbb{E}\left(z_t|\mathcal{F}_t^{LI}\right) < 0$ for any $\eta_{t-1} \leq \eta_{t-1}^{+}$; the probability that $\eta_{t-1} \leq \eta_{t-1}^{+}$ is strictly positive.

As a short summary, Figure 1 provides a graphical illustration of the formal results obtained in Propositions 1 and 2. A shock occurs in period $t = 4$. In Subfigures (a) and (b) we observe the conditional expectation of the limited informed agent and the output of the informed and the limited informed agents given the assumptions of the base-line model presented in Section 2. We attain convergence to the true parameter $\theta = 1$ and diminishing differences between the quantities $x_t$ and $x_t^{FI}$. In Subfigures (c) and (d) the Assumptions 1-3 hold. This results in negative $\mathbb{E}\left(z_t|\mathcal{F}_t^{LI}\right)$ and $\mathbb{E}\left(\theta|\mathcal{F}_t^{LI}\right)$ in period $t = 5$. While the fully informed agents start production after some periods (Case C2), the limited informed agents exit (Case C4) in Subfigure (d). By applying (14), $\mathbb{E}\left(z_t|\mathcal{F}_t^{LI}\right)$ converges to $\mathbb{E}\left(\theta|\mathcal{F}_t^{LI}\right)$ after exit as observed in Subfigure (c).

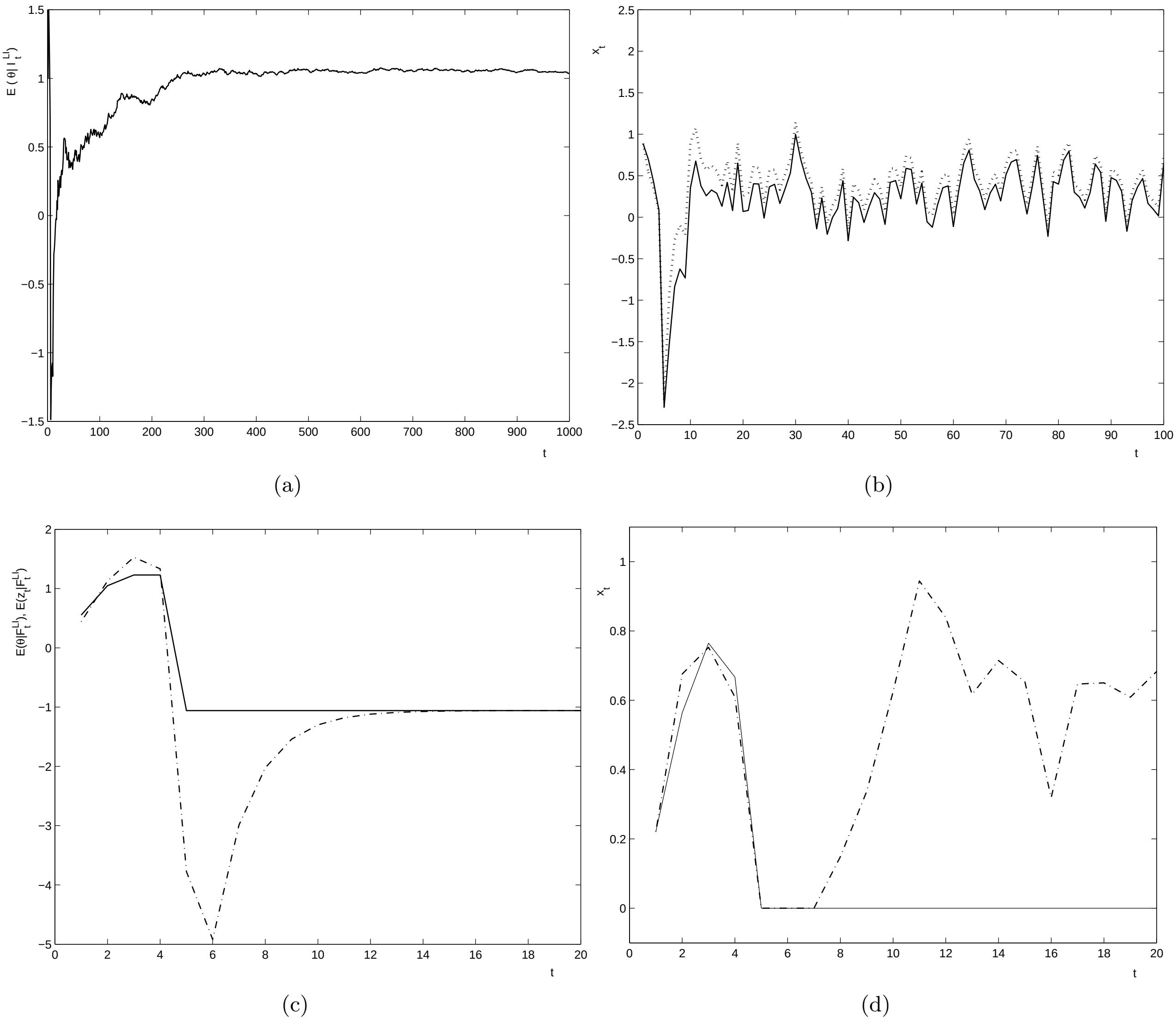


**Figure 1: Quantities, Conditional Expectations and Convergence.** This figure plots time series of the quantities produced and conditional expectations. Parameters set to: $\zeta = 0.5$, $\theta = \theta_0 = 1$, $\sigma_\eta^2 = \sigma_0^2 = 1$, $\lambda = \beta = 1$. A shock occurs in period $t = 4$. Subfigures (a) and (b) present representative output from the baseline model of Section 2: (a) plots $\mathbb{E}\left(\theta|\mathcal{I}_t^{LI}\right)$ for $t = 1, \ldots, 1000$. (b) plots $x_t$ (solid line) and $x_t^{FI}$ (dotted line) for the first 100 periods. Subfigures (c) and (d) present a non-convergent path for the setting presented in this section. (c) Conditional expectation $\mathbb{E}\left(z_t|\mathcal{F}_t^{LI}\right)$ (dashed-solid line) and $\mathbb{E}\left(\theta|\mathcal{F}_t^{LI}\right)$ (solid line) for the limited information case. (d) output $x_t$ (solid line) and $x_t^{FI}$ (dashed-solid line).

**Remark 4.** Assumption 1, $x_{it} \geq 0$, could also be replaced by $x_{it} \geq \bar{x}$, where $\bar{x}$ could be a minimum production level or a short selling constraint. For restrictions like this the results from this section can be adapted.

# 4 Bounded Support

An assumption that avoids the above implications is as follows:

**Assumption 4.** The support of $\eta_t$ is a proper subset of $[\underline{\eta}, \bar{\eta}]$ and the parameter $\theta > 0$. W.l.g. $\mathbb{E}(\eta_t) = 0$. $Supp_\eta$ is the support of $\eta_t$.

As in Section 3: (i) If $x_t > 0$, then $p_t = z_t - \beta x_t$ and $y_t = z_t$. (ii) If $x_t = 0$, $p_t$ and $z_t$ are missing values by Assumption 2. The first order condition (15) continues to hold. $z_0 = \theta$ and (3) imply:

$$z_t = \theta(1 - \zeta) + \zeta z_{t-1} + \eta_t = \theta + \sum_{s=0}^{t-1} \zeta^s \eta_{t-s} \ . \tag{16}$$

With $|\zeta| < 1$ and bounded support, the absolute value of the last term in (16) remains smaller than $\frac{1}{1-|\zeta|} max\{\bar{\eta}, |\underline{\eta}|\} =: \eta_{max}$. Therefore, with $|\zeta| < 1$, $z_0 = \theta$ and $\eta \in [\underline{\eta}, \bar{\eta}]$, we get the lower and upper bounds

$$\underline{z} \quad = \quad \theta - \eta_{max} \text{ and } \bar{z} = \theta + \eta_{max} \ . \tag{17}$$

To derive a positive $z_t$ we have to restrict the support of $\eta_t$ such that $\underline{z} > 0$. This yields $\eta_{max} < \theta(1-|\zeta|)$. The requirement $<$ comes from $z_t > 0$ with an arbitrary distribution of $\eta_t$ (with bounded support), i.e. also atoms at the borders are allowed. As long as the distribution of $\eta_t$ is absolutely continuous, $<$ can be replaced by $\leq$. This yields:

**Corollary 2.** Suppose that Assumptions 1-4 hold and the support of $\eta_t$ is a proper subset of $[-(1 - |\zeta|)\theta, (1-|\zeta|)\theta]$. In addition firms are equipped with a Bayesian updating scheme such that $\mathbb{E}(\theta|\mathcal{F}_t^{LI}) \to \theta$ as $t \to \infty$ ($a.s.$ and $L^2$) and $\mathbb{E}(\theta|\mathcal{F}_t^{LI}) \in (\underline{\theta}, \bar{\theta})$ for all $t \geq 0$; $\underline{\theta} = \theta - \frac{1}{1-|\zeta|} max\{\bar{\eta}, |\underline{\eta}|\}$ and $\bar{\theta} = \theta + \frac{1}{1-|\zeta|} max\{\bar{\eta}, |\underline{\eta}|\}$.

Then $\mathbb{E}(\theta|\mathcal{F}_t^{(.)}) > 0$ and $\mathbb{E}(z_t|\mathcal{F}_t^{(.)}) > 0$; $x_t > 0$ for all $t$ by the first order condition (15) (i.e. only C1 of Proposition 2 is possible). The converge results of Proposition 1[part (i)] continue to hold.

Bayesian learning can be replaced by the weaker assumption that agents use an estimator $\hat{\theta}_t$, with $\hat{\theta}_t \to \theta$, where $\hat{\theta}_t$ and the forecasts of $z_t$ remain in $[\underline{z}, \bar{z}]$.

Corollary 2 demands for positive $z_t$, consistency with respect to the support and consistency of the Bayes estimate. That is to say, a bounded support of the random variable, such that $z_t > 0$, is not sufficient for convergence. The following examples should shed some light on these issues:

**Example 1** (Truncated normal with miss-specified upper and lower bounds)**.** Suppose that $\theta > 0$, and $\varepsilon_t$ follows a truncated normal distribution with lower bound $\underline{\varepsilon}$ and upper bound $\bar{\varepsilon}$, mean parameter $\nu$ and variance parameter $\sigma^2_{TN}$ resulting in $z_t > 0$ (for the truncated normal we refer the reader to Appendix A). Truncated normal priors are assumed. In addition assume that $\zeta = 0$, such that $\mathbb{E}(z_t|\mathcal{F}_t^{LI}) < 0$ is sufficient for exit. Take (18) of Appendix A and a parameter $\nu$ such that $\mathbb{E}_{f_{TN}(\nu,\sigma^2_{TN},\underline{\varepsilon},\bar{\varepsilon})}(z_t) = \theta > 0$. These assumptions result in $\underline{z} = \underline{\varepsilon}$ and $\bar{z} = \bar{\varepsilon}$; $\underline{\eta}$ and $\bar{\eta}$ can be obtained by means of $\underline{\eta} = \underline{\varepsilon} - \theta$ and $\bar{\eta} = \bar{\varepsilon} - \theta$.

Suppose that firms know that the data are generated from a truncated normal but assume a lower bound $\underline{\tilde{\varepsilon}} < \underline{\varepsilon}$ and an upper bound $\bar{\tilde{\varepsilon}}$; these bounds are fixed and the distribution can be asymmetric. By (18) we find upper and lower bounds such that $0 < \mathbb{E}_{f_{TN}(\nu,\sigma^2_{TN},\underline{\tilde{\varepsilon}},\bar{\tilde{\varepsilon}})}(\varepsilon_t) < \theta$. Now suppose that the firms apply Rodriguez-Yam et al. (2004) (as in Example 9) to estimate $\nu$. Then a sufficiently small $z_t > 0$ can result in an estimate $\hat{\nu}_t$ such that $\mathbb{E}_{f_{TN}(\hat{\nu}_t,\sigma^2_{TN},\underline{\tilde{\varepsilon}},\bar{\tilde{\varepsilon}})}(\varepsilon_t) = \mathbb{E}(z_t|\mathcal{I}_t^{LI}) < 0$ where firms exit.

**Example 2** (Truncated normal, non-regular case)**.** The miss-specification of the econometric model in Example 1 can be repaired by firms estimating the bounds of the truncated normal. Example 9 in Appendix B demonstrates that the estimates are still consistent if $(z_t)$ is observed and the upper and lower bounds are unknown parameters. The bounds converge to the true bounds as $t \to \infty$. This does not imply $\mathbb{E}(z_t|\mathcal{F}_t^{LI}) > 0$ almost surely for all $t$. E.g. if the firms put relatively strong priors in the neighborhood of $\bar{\tilde{\varepsilon}}$ and $\underline{\tilde{\varepsilon}}$ assumed in Example 1 convergence need not take place.

**Example 3** (Truncated normal with correctly specified upper and lower bounds)**.** Given the assumptions of Example 1, but the lower and upper bounds assumed correspond to the true parameters. Then $\mathbb{E}(z_t|\mathcal{F}_t^{LI}) > 0$ by applying Rodriguez-Yam et al. (2004) and (7).

The above examples demonstrate why bounded support of $(z_t)$ is not sufficient for convergence. In Example 1 the firms apply a prior and a likelihood with inconsistent bounds. That is to say, firms know or guess the correct class of distributions, but use a miss-specified econometric model with respect to the range of the random variable. Here also the estimates in the baseline case would be inconsistent. Example 2 repairs this drawback by estimating the bounds. Although the posterior will converge to the true parameter $\psi$ if $(z_t)_{t=0}^{\infty}$ is observed, this does not imply that the bounds for exit cannot be hit. Example 3 works with correctly specified bounds of the distributions. This is a strong assumption since firms exactly know the worst and the best shock that could happen.

**Remark 5.** Corollary 2 does not guarantee non-negative prices. The inverse demand function (1) and the first order condition (15) result in $p_t = z_t - \frac{\beta}{\beta+\lambda}\mathbb{E}(z_t|\mathcal{F}_t^{LI})$. Even with the assumptions of Corollary 2 where $\theta > 0$ and $\eta \in (\underline{\eta}, \bar{\eta})$, the price can become negative if $z_t$ is small but the expectation is large. With forecasts in $[\underline{z}, \bar{z}]$, by a simple plug in of the upper and lower bounds of $z_t$ we get $p_t \geq 0$ if $\underline{z} \geq \bar{z}\frac{\beta}{\beta+\lambda}$.

**Remark 6.** Non-negative prices and quantities can also be obtained in a different ways. E.g. Adam and Marcet (2010) work with *log* returns, while Vives (2008)[p. 271] propose a model with isoelastic cost and inverse demand structure. By these assumptions the authors restrict to Case 1 of Proposition 2. With isoelastic demand an $x_t \to 0$ forces prices to go to $+\infty$ for any signal. This guarantees an interior solution with $x_t > 0$. A further opportunity is to transform $z_t$ by a function $g(.)$ such that $g(z_t)$ fulfills the requirements of Corollary 2. For example the logistic function, the arcus tangens function, etc. can be used in this case. In addition Bayesian learning results in non-negative forecasts.

*Robustness*: Based on these last examples and Remark 6, we might expect that with bounded support consistency and convergence take place under fairly mild conditions. "Fairly mild" is a matter of taste. Even with innovations on bounded support and $\theta \in \eta \in (\underline{\theta}, \bar{\theta})$ we can use a Dirichlet mixtures as already discussed in the last paragraph of Section 2 and in Appendix B. Similar to the asymmetric normal distribution constructed in Appendix A, a truncated asymmetric normal distribution can be constructed in the same way:

**Example 4** (Mixture of Truncated Normals)**.** Take the upper and lower bounds $\underline{\varepsilon}$ and $\bar{\varepsilon}$ from the Example 3 such that $z_t > 0$ and assume that these bounds are known by all agents. Let $\varepsilon_t$ follow a truncated normal with parameter $\nu$ for irrational $\nu$, while $\varepsilon_t$ is asymmetric truncated normal of Fernandez and Steel (1998) type with some asymmetry parameter $\gamma \neq 1$ (see Appendix A). Once again the Bayes estimate of $\nu$ is given Rodriguez-Yam et al. (2004) (see Example 9 in the Appendix B). This estimate converges to the true parameter for all irrational $\nu$ while for rational $\nu$ we get inconsistent estimates.[7]

In Example 4 the requirements of Case C1 in Proposition 2 are satisfied all the time, i.e. we have excluded shut down. However, also under these circumstances the Bayes estimate need not be consistent and convergence of the limited to the full information case need not take place for all parameter values of the parameter space. This implies, that we can also construct examples with inconsistent estimates although shut down is not observed. Similar examples can be constructed with other distributions where the distribution of the noise is from a different class for rational and irrational elements. Therefore, even if we replace the linear demand structure by a log-linear one to force prices and quantities to be positive (as suggested by Remark 6), inconsistency of the Bayes estimate can still arise. Despite the fact that we can make the model more realistic by different assumptions to make prices and quantities non-negative, learning the true parameter still remains an issue in all these cases.

## 5 Conclusions

By simple considerations we observed that by eliminating negative quantities but maintaining the linear normal demand structure, the converge results derived in Jun and Vives (1996) and Vives (2008) need not hold. By including a simple shutdown condition - even in the limit - the quantities provided in a *full information* economy need not agree with the the quantities produced in the *limited information* setting. To rescue the optimistic convergence results bounded support of the noise term and regularity conditions on learning are required.

[7]Note that consistency of an estimate means that for *all* elements of the parameter space the posterior converges to a point mass at the true parameter (see Appendix B).

## A The Truncated Normal and the Asymmetric Normal Distribution

Consider a truncated normal random variable $X$ with density $f_{TN}(x;\nu,\sigma^2_{TN},\underline{\varepsilon},\bar{\varepsilon})$ (see e.g. Paolella (2007)). The expected value of a truncated normal random variable $X$ with location and scale parameters $\nu,\sigma^2_{TN}$ and bounds $\underline{\varepsilon}<\bar{\varepsilon}$ is

$$\mathbb{E}(X)=\nu+\frac{f_{SN}((\underline{\varepsilon}-\nu)/\sigma_{TN})-f_{SN}((\bar{\varepsilon}-\nu)/\sigma_{TN})}{F_{SN}((\bar{\varepsilon}-\nu)/\sigma_{TN})-F_{SN}((\underline{\varepsilon}-\nu)/\sigma_{TN})}\sigma_{TN}\ . \tag{18}$$

$f_{SN}$ and $F_{SN}$ are a standard normal density and distribution function.

Based on Fernandez and Steel (1998) an asymmetric distribution with density $g(x)$ can be constructed from a symmetric distribution with density $f(x)$ as follows (see e.g. also Paolella (2007)):

$$g_\gamma(x) \;=\; 2\frac{\zeta}{1+\zeta^2}\left(f(x/\gamma)\mathbf{1}_{x\geq 0}+f(\gamma x)\mathbf{1}_{x<0}\right)\ . \tag{19}$$

The parameter $\gamma>0$ controls for the degree of asymmetry. For $\gamma=1$, the distribution is symmetric such that $g(x)=f(x)$. The moments of $X$ are given by:

$$\mathbb{E}(X^r|\gamma) \;=\; \frac{\zeta^{r+1}+(-1)^r/\zeta^{r+1}}{\zeta+1/\zeta}2\mathbb{E}\left(X^r|X>0,\gamma=0\right)\ . \tag{20}$$

By using the standard normal density $f_{SN}$ we get an asymmetric normal distribution with $\mathbb{E}\left(X\right)\neq 0$ for $\gamma\neq 1$. For the normal distribution $\mathbb{E}\left(X^r|X>0,\gamma=0\right)$ can be derived as follows: Given the stardard normal density $f_{SN}(x)$ we observe that $f'_{SN}(x)=(-x)f_{SN}(x)$. By partial integration we observe that $\mathbb{E}\left(X^r|X>0,\gamma=0\right)=\int_0^\infty x^r f_{SN}(x)dx=\frac{1}{r+1}x^{r+1}f_{SN}(x)|_0^\infty-\frac{1}{r-1}\int_0^\infty x^{r+1}(-x)f_{SN}(x)dx=\frac{1}{r+1}\int_0^\infty x^{r+2}f_{SN}(x)dx$ such $\mathbb{E}\left(X^r|X>0,\gamma=0\right)=\frac{1}{r+1}\mathbb{E}\left(X^{r+2}|X>0,\gamma=0\right)$. Since $f'_N(x)=(-x)f_{SN}(x)$ we get $\int_0^\infty f'_{SN}(x)dx=\int_0^\infty(-x)f_{SN}(x)dx$ yielding $\mathbb{E}\left(X^1|X>0,\gamma=0\right)=\frac{1}{\sqrt{2\pi}}$, while $\mathbb{E}\left(X^2|X>0,\gamma=0\right)=\frac{1}{2}$ for a standard normal random variable. By the above recursive relationship of the moments we get $\mathbb{E}\left(X^r|X>0,\gamma=0\right)$ for the standard normal. Some algebra yields

$\mathbb{E}\left(X^r|X>0,\gamma=0\right) = \frac{r!}{2^{r/2}(r/2)!}$ for even $r$ and $\mathbb{E}\left(X^r|X>0,\gamma=0\right) = 2^{(r-1)/2}((r-1)/2)!\frac{1}{\sqrt{2\pi}}$ with $r$ odd.

Given a standardized skewed normal random variable $X$, we get as usual a non-standardized normal random variable $Y$ by means of $Y=\theta+\sigma X$, where $\theta$ and $\sigma^2$ are the mean and the variance of $Y$ with $\gamma=1$. By means of (19) we can also construct an asymmetric truncated normal distribution by using $f_{TN}(.)$ instead of $f(.)$.

# B Bayesian Learning and Consistency

This section reviews Bayesian learning and presents examples with non-Gaussian innovations. Regarding convergence, we refer the reader to Ghosal et al. (2000) and the web-note of Shalizi (2010), describing the problems and providing literature on this topic.

Let us start with a prior $\pi$ on $\Psi$. $\Psi$ is the parameter space, the elements are $\psi$. Given, the distribution of $x^t=(x_1,\ldots,x_t)$, $f(x^t|\psi)$, the posterior can be derived by means of the Bayes theorem

$$\pi(\psi|x^t) = \frac{f(x^t|\psi)\pi(\psi)}{\int_\Psi f(x^t|\psi)\pi(d\psi)} \ . \tag{21}$$

Only for a small number of applications $\pi(\psi|x^t)$ can be derived analytically. Bayesian simulation methods to simulate from the posterior $\pi(\psi|x^t)$ is described in Appendix C.

A more detailed picture is obtained by following Diaconis and Freedman (1986) and Blume and Easley (1993): The parameter space $\Psi$ is a Borel subset of a complete separable metric space, the elements are $\psi$. Suppose that $\pi$ is non-degenerated and $\pi>0$ for any Borel subset containing the true parameter generating the data. $Q_\psi$ is a probability measure on Borel subsets of the set of data $X$, $x\in X$. $Q_\psi^\infty$ is the product measure on $X^\infty$. The map $\psi\to Q_\psi$ is one to one and Borel, the point mass at $\psi$ is denoted by $\delta_\psi$. The joint distribution $P_\pi(A\times B)=\int_A Q_\psi^\infty(B)\pi(d\psi)$, where $A$ and $B$ are Borel in $\Psi$ and $X^\infty$, respectively. For a Borel subset $H$ of $\Psi\times X^\infty$ we get $P_\pi(H)=\int_\Psi(\delta_\psi\times Q_\psi^\infty)(H)\pi(d\psi)$. $\delta_\psi$ is the point mass at $\psi$. Then the posterior $\pi(\psi|B)$ - also abbreviated by $\pi(\psi|x^\infty)$ - is a version of the $P_\pi$ law of $\psi$ (Consistency does not only depend on the prior but also on the version of the posterior we choose; for more details see Diaconis and Freedman (1986)[Example A.1]). For a measurable set $B^t$, this results in the posterior $\pi(\psi|B^t)$, or $\pi(\psi|x^t)$ in the other notation used in (21).

For convergence issues in this context usually the weak-star topology is applied: Consider the probability measures $\mu_t$ and $\mu$ on $\Psi$: $\mu_t \to \mu$ weak star if and only if $\int f d\mu_t \to \int f d\mu$ for all bounded continuous functions $f$ on $\Psi$. $\mu_t \to \delta_\psi$ if and only if $\mu_t(U) \to 1$ for every neighborhood $U$ of $\psi$. $\delta_\psi$ is the Dirac point mass at $\psi$. By Doob's theorem $\pi(\psi|x^t) \to \delta_\psi$ in the weak star topology $P_\pi$ almost surely expect possibly on a set of $\pi$ measure zero (see Diaconis and Freedman (1986)[Corollary A.2], Ghosal et al. (2000) and Blume and Easley (1993)[Theorem 1]). That is to say that the Bayesian is subjectively certain to converge to the truth, or in other terms, that for almost all $H \in \Psi \times X^\infty$ the posterior is consistent. However, in frequentist terms: suppose that some $\psi \in \Psi$ generates the data, we do not know whether the estimator is consistent. Lijoi et al. (2007) recently derived an extension of Doob's convergence result for stationary data.

*Consistency at* $\psi$ means that $\pi(U|x^t) \to 1$ (in the probability measure $P_\psi$) for any neighborhood $U$ of $\psi$. *Consistency* implies that the posterior converges to the true parameter $\psi$; that is to say we have consistency at $\psi$ for all elements of the parameter space (see e.g. Diaconis and Freedman (1986) or Pötscher and Prucha (1997)). Doob's theorem states that for any prior $\pi(\psi)$ the posterior is consistent at any $\psi \in \Psi$ except possibly on a set of $\pi$ measure zero (see Ghosal et al. (2000)). This does not guarantee convergence to the true parameter for *all* $\psi \in \Psi$ as required. Even worse, given a set of priors $\mathcal{M}$, Diaconis and Freedman (1986) have shown that the set of all pairs $(\psi, \pi) \in \Psi \times \mathcal{M}$ for which the posterior based on $\pi$ is consistent is a meager set (of Baire category 1; here the topology of weak convergence has been used). The original result has been shown on a countable parameter space, while Feldman (1991) extended this result to a complete and separable, non-finite parameter space.

**Remark 7.** By the latter result we observe that there are many "bad priors", the set of "good priors" is topologically small. On the other hand if the statistician - in our case the firm - is "certain about the prior $\pi(\psi)$", Doob's theorem tells us that the set where inconsistency might arise is small in a measure theoretic sense. The key aspect is that the firm has to choose a good prior as e.g. in the benchmark model. However, choosing a good prior is in general a strong assumption on the capabilities of the economic agents.

Statistical literature provides examples where Bayesian learning is inconsistent; here the reader is referred e.g. to Diaconis and Freedman (1986) and the literature cited there. Especially, when non-parametric Bayesian techniques are used, inconsistency becomes a serious problem. Consistency for non-parametric Bayesian models has been investigated more recently in Barron et al. (1999), Walker (2004), Ghosal and van der Vaart (2007) and Rousseau et al. (2010). For sufficient conditions for consistency and (counter-)examples in the non-parametric case we refer the reader to Jang et al. (2010). Based on Christensen (2009) we get the following example:

**Example 5** (Inconsistency, Application of Christensen (2009))**.** Given the assumptions of Section 2, but $\eta_t \sim \mathcal{N}(\theta, \sigma_\eta^2)$ for irrational $\theta$, while $\eta_t$ is Cauchy with location parameter $\theta$ and variance parameter $\sigma_e^2$ for rational $\theta$. By sticking to the normal prior $\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$, the posterior $\pi(\theta|z^t) = \pi(\theta|\mathcal{I}_{t+1}^{LI})$ is still given by (10), where $\theta \sim \mathcal{N}(a_t, A_t)$. For irrational $\theta$ this estimate is consistent, while for rational $\theta$ we get inconsistent estimates.

In this example we get inconsistency on a set of prior probability 0, i.e. this set is small in a measure theoretic sense. However, we get inconsistent estimates on a dense set. A set is small in a topological sense (meager set or set of the first Baire category), if it can be expressed as the countable union of nowhere dense sets. Using this topological concept, the set of consistent estimates is meager in this example. A more heuristic interpretation of this results is as follows: In any case an agent can only report rationals/key in rational numbers into a computer. However, with any of these rational numbers the estimate is not consistent all the time. In addition if the sampling distribution of the unknown parameter $\theta$ puts positive probability mass on the rational numbers, the problem becomes even worse.

**Remark 8.** From (10) we observe that with Gaussian noise the estimator of $\theta$ is a convex combination of the sample mean and a term arising from the prior which becomes neglectable if $t$ becomes larger. For independent Gaussian noise the strong law of large number holds such that the sample mean converges to $\theta$. For a Cauchy distribution this is not the case. In Example 5 this is the main driving force for the inconsistency of the estimator on rational $\theta$.

If distributions with non-finite expectation do not seem very realistic to the reader, we can think of an economy where estimates of $\theta$ and $\sigma_\eta^2$ enter into equilibrium quantities. Suppose that $\theta$ is fixed while

$\sigma_\eta^2$ is unknown. The expected value should be $\theta$ for all $\sigma_\eta^2$. Assume a normal distribution for irrational $\sigma_\eta^2 > 0$ and a distribution where the second moment does not exist for rational $\sigma_\eta^2 > 0$ (e.g. a student t-distribution with $1 < \nu \leq 2$ degrees of freedom, a distribution symmetric with respect to $\theta$ where $\mathbb{E}(X) = \theta$ and $\mathbb{E}(X^2)$ does not exist). In this case the posterior based on the Gaussian distribution is an inverse Gamma distribution (see Robert (1994)[Chapter 4] or Frühwirth-Schnatter (2006)), where the sample variance determines the *scale parameter* of this distribution. By the same argument as above we get inconsistent estimates if the true $\sigma_\eta^2$ is a rational number.

**Example 6** (Inconsistency, Variation of Example 5)**.** Similar to Example 5 we can also construct Dirichlet mixtures with finite expectation. Let $\eta_t \sim \mathcal{N}(\theta, \sigma_\eta^2)$ for irrational $\theta$, while $\eta_t$ is asymmetric normal of Fernandez and Steel (1998) type (see Appendix A). $\pi(\theta|z^t) = \pi(\theta|\mathcal{I}_{t+1}^{LI})$ is once again given by (10). For irrational $\theta$ this estimate is consistent such that $a_t \to \theta$, while for rational $a_t \to \theta + \mathbb{E}(\eta_t)$ by the law of large numbers. This implies $\lim a_t \neq \theta$ for the asymmetric case. A similar example can be constructed with mixtures of asymmetric truncated normal distributions, etc.

One of the usual procedures to check for convergence to the true parameter for a particular model is to show that the probability of non-convergence to the true parameter value $\psi$ goes to zero (see also Blume and Easley (1993)[Corollary 2.1]) or to work with so called *consistency hypothesis tests* as discussed and presented in Ghosal et al. (1994), Ghosal et al. (1995a), Ghosal (2000) Ghosal et al. (2000), Ghosal and Tang (2006), Ghosal and van der Vaart (2007) and Walker (2004). Section 2 investigated a regular problem, where the support of the stochastic noise term does not change with the parameter $\psi$. For regular problems the result derived by Schwartz (1965) can often be applied to check for consistency. Further regular problems will be presented in Examples 8 and 9. Results for non-regular problems have also been investigated in Ghosal et al. (1994), Ghosal et al. (1995a), Chernozhukov and Hong (2004) and Hirano and Porter (2003). The following theorem provides sufficient conditions for almost sure convergence to the true parameter:

**Proposition 3.** *[Proposition 1 in Ghosal et al. (1995b); based on Ibragimov and Has'minskii (1981)] Consider a prior $\pi$, $\Psi \in \mathrm{R}^d$ and $U_t = \varphi_t^{-1}(\Psi - \psi)$ for some normalizing factor $\varphi_t$. The likelihood ratio* *process is defined as*

$$L_t(u) = \frac{f(x^t|\psi + \varphi_n u)}{f(x^t|\psi)} \ , \ u \in U_t.$$

*Assume that the conditions*

*IH1 For some $M > 0$, $m \geq 0$ and $\alpha > 0$,*

$$\mathbb{E}_\psi \left( \sqrt{L_t(u_1)} - \sqrt{L_t(u_2)} \right)^2 \leq M(1 + R^m) \parallel u_1 - u_2 \parallel^\alpha \ ,$$

*for all $u_1, u_2 \in U_t$ satisfying $\parallel u_1 \parallel \leq R$ and $\parallel u_2 \parallel \leq R$.*

*IH2 For all $u \in U_t$,*

$$\mathbb{E}_\psi \left( \sqrt{L_t(u)} \right) \leq \exp\left(-g_t \parallel u \parallel\right) \ ,$$

*where $\{g_t\}$ is a sequence of real valued functions on $[0, \infty)$ satisfying the following: (a) for a fixed $t \geq 1$, $g_t(y) \to \infty$ as $y \to \infty$ and (b) for any $N > 0$, $\lim_{y,t\to\infty} y^N \exp\left(-g_t(y)\right) = 0$.*

*GGS For some $s > 0$, $\sum_{t=1}^{\infty} \parallel \varphi_t \parallel^s < \infty$.*

*are satisfied, then almost sure convergence of the posterior to the true parameter is attained.*

Usually $\varphi_t = 1/\sqrt{t}$ is applied. When only conditions IH1 and IH2 are satisfied, we get convergence in probability. It is worth noting that Proposition 3 does neither require *iid* data nor a regular problem; only the existence of the densities (likelihood) $f(x^t|\theta)$ is assumed. Proposition 3 also tells us that the prior has to be strictly positive around the true parameter value and that it is only allowed to grow at most like a polynomial. $L^p$ convergence generally does not follow from almost sure convergence. If almost sure convergence is observed and the $p$-th moment of the sequence of random variables is uniformly integrable, then $L^p$ convergence is attained e.g. by means of by Lebesgue's dominated convergence theorem (see e.g. Billingsley (1986)[p. 213] or Klenke (2008)[p. 140]).[8]

[8] A further issue is the limit distribution of the posterior around the true parameter value. For finite dimensional parameter spaces the Bernstein-von Mises Theorem (see e.g. LeCam and Yang (1990), Lehmann (1991), Freedman (1999), Bickel and Doksum (2001)[p. 339]) states that the Bayesian posterior around its mean is close to the asymptotic distribution of the maximum likelihood estimate around the true parameter value. In this case consistency of the maximum likelihood estimate implies consistent Bayes estimates. Also here Ghosal et al. (1994), Ghosal et al. (1995a) and Ghosal and van der Vaart (2007) provide useful tools to derive the limit distribution of the Bayes estimate.

In Section 2, the parameter $\psi$ is $\theta$, $\Psi$ is the real line. (10) provides the posterior $\pi(\theta|\mathcal{I}_t^{LI})$ in closed form. Consistency $\pi(\theta|\mathcal{I}_t^{LI}) \to \delta_\theta$ can be checked directly by taking limits. Alternatively, consistency follows from Blume and Easley (1993)[Theorem 2.2]. Learning the mean of a normal random variable is also presented as an example there. $L^p$ convergence follows from the properties of the normal distribution. The conditional expectation of $z_t$ was given by a convex combination of the parameter $\theta$ and the former realization $z_{t-1}$. Convergence results on the parameter $\theta$ automatically carry over to convergence of the quantities. Generally, this need not be as simple, as demonstrated in the following example:

**Example 7.** Assume that $\eta_t$ is Cauchy. Suppose that the posterior of the location parameter $\theta$ converges to the true parameter $\theta$. In this setting the parameter $\psi = \theta$ has been learned. In contrast to learning the expectation of $\eta_t$ does not exist. As a consequence the conditional expectation of $z_t$ does not exist (neither in the full information nor in the limited information setting) and no convergence results on $(x_t)$ can be obtained.

Example 7 demonstrates that consistent estimates are in general not sufficient for convergence (even with $|\zeta| < 1$). Hence, we can additionally require uniform integrability of the random variables $z_t$ or $\Delta^\zeta z_t$. Based on this discussion Proposition 1 can be extended as follows:

**Corollary 3.** [Extension of Proposition 1]
Suppose $\varepsilon^t \sim g(\varepsilon^t|\psi)$, $\mathbb{E}(|\varepsilon_t|^p) < \infty$ for all $t$ and some $p \geq 1$. In addition normalize the noise term such that $\mathbb{E}(\Delta^\zeta z_t) = \mathbb{E}(\varepsilon) = \theta(1-\zeta)$. The unknown parameters are $\psi$, $\Psi \in \mathrm{R}^d$.

If $|\zeta| \neq 1$ and the consistency conditions of Proposition 3 are fulfilled, then the posterior converges to the true parameter $\psi$ almost surely. If, in addition, $|\zeta| \neq 1$ and $|\mathbb{E}\left(\psi|\mathcal{I}_t^{LI}\right)|^p$ is uniformly integrable for $p \geq 1$ then $\mathbb{E}\left(\psi|\mathcal{I}_t^{LI}\right)$ converges to $\psi$ in $L^p$. For $|\zeta| < 1$ and $\mathbb{E}\left(\psi|\mathcal{I}_t^{LI}\right) \to \psi$ almost surely and $L^p$ then $x_t \to x_t^{FI}$ almost surely and in $L^p$. For $\zeta \in \mathrm{R}$, $\zeta \neq 1$ and $\mathbb{E}\left(\psi|\mathcal{I}_t^{LI}\right) \to \psi$ almost surely in in $L^p$, $\Delta^\zeta x_t$ converges to $\Delta^\zeta x_t^{FI}$ a.s. and $L^p$. For $\zeta = 1$, $x_t = x_t^{FI}$ for $t \geq 2$.

The following examples illustrate the implications of Corollary 3. In the first example we replace the normally distributed normal noise term with a student $t$-distributed noise term with $\nu$ degrees of freedom. The second example investigates the truncated normal case.

**Example 8.** Assume $\zeta \neq 1$ and $\Delta^{\zeta} z_t = (1-\zeta)\theta + \eta_t$, where $\eta_t$ is student $t$ centered around zero with $1 < \kappa \leq 2$ degrees of freedom. The variance parameter $\sigma_\eta^2$ is fixed. For the parameter $\theta$ the agents apply a student $t$ prior with $\nu$ degrees of freedom, mean parameter $m_0$ and variance parameter $\sigma_\eta^2/M_0$. W.l.g. set $\sigma_\eta^2 = 1$. Then the posterior of the unknown parameter $\theta$ is student-t with variance parameter $M_t = 1/(t(1-\zeta)^2 + M_0)$ and mean parameter $m_t = \frac{(1-\zeta)\sum_{s=1}^{t} \Delta^{\zeta} z_s + M_0 m_0}{t(1-\zeta)^2 + M_0}$ (see Harrison and West (1997)[p. 422]). In the same way as with the baseline model we get $\mathbb{E}\left(\theta | \mathcal{I}_t^{LI}\right) = m_{t-1}$. $m_t$ converges to $\theta$ (this could also be directly observed by applying Etemadi's version of the strong law of large number Klenke (2008)[p. 112]). Here only $L^{\kappa'}$ converge, with $\kappa' < \kappa$, follows from the dominated convergence theorem.

**Example 9.** Given the assumptions of Section 2: $\Delta z_t = \varepsilon_t$. $\varepsilon_t$ follows a truncated normal distribution with lower bound $\underline{\varepsilon}_t$ and upper bound $\bar{\varepsilon}_t$. The mean parameter is $\nu$, the variance parameter $\sigma_{TN}^2$ is fixed, $\nu \in [\underline{\varepsilon}_t, \bar{\varepsilon}_t]$. $\nu$ is such that $\mathbb{E}(\varepsilon_t) = \theta(1-\zeta)$. Suppose a normal or a truncated normal prior. With the variance parameter and the bounds fixed, we verified that the conditions of Proposition 3 hold for the unknown parameter $\nu$. Therefore, we attain almost sure converge to the true parameter value $\nu$. Although $L^2$ convergence generally depends on the prior $\pi$ used, if agents assume a (truncated) normal prior then $L^2$ convergence is obtained.

Suppose in addition that agents assume a truncated normal prior with parameters $\nu_{TN0}$ and $\sigma_{TN0}^2$ and the *same* lower and upper bounds. Rodriguez-Yam et al. (2004) have shown that the posterior of $\nu$ based on the data $\Delta^{\zeta} z^t$ is a truncated normal distribution with mean parameter $a_{TNt} = \frac{(1/\sigma_\varepsilon^2)\sum_{s=1}^{t} \Delta^{\zeta} z_s + \nu_{TN0}/\sigma_{TN0}^2}{(1/\sigma_{TN}^2)t + 1/\sigma_{TN0}^2}$ and variance parameter $A_{TNt} = \left((1/\sigma_{TN}^2)t + 1/\sigma_{TN0}^2\right)^{-1}$.[9]

In Examples 8 and 9 we investigated regular problems, i.e. problems where the support of the stochastic noise term does not change with the parameter $\Psi$. In Example 9 knowledge of the upper and the lower bounds is a strong assumption. By applying Proposition 3 we derive the following result:

**Example 10** (Application of Ghosal et al. (1995a)). Given the assumptions of Example 9, but the lower and the upper bounds of the truncated normal distribution are unknown, i.e. the true parameter

[9] In the baseline model and in this example it is not difficult to verify that the estimates of $\psi = \{\nu, \sigma_{TN}^2\}$ are also consistent, i.e. also the variance parameter can be estimated consistently. Simulation methods as described in Appendix C have to be applied to derive samples from the posterior.

generating the data is $\psi = \{\nu, \underline{\varepsilon}, \bar{\varepsilon}\}$. The parameter space $\Psi \subset \mathrm{R}^3$, with $\tilde{\underline{\varepsilon}} < \bar{\tilde{\varepsilon}} \in \Upsilon$. Apply a normal prior to $\nu$. Either (i) $\Upsilon$ is a compact subset of $\mathrm{R}^2$, $\Upsilon$ is assumed to be "sufficiently large", and a uniform prior is applied for any pairs $\tilde{\underline{\varepsilon}}, \bar{\tilde{\varepsilon}} \in \Upsilon$; or (ii) normal priors are applied to $\tilde{\underline{\varepsilon}}, \bar{\tilde{\varepsilon}}$. $\Upsilon \subset \Psi$ With these priors we verified that the conditions of Proposition 3 hold. This implies that the posterior converges almost surely to the true parameter.

# C Bayesian Parameter Estimation

In most cases $\pi(\psi|x^t)$ cannot be derived analytically. Fortunately, statistical literature has developed a bulk of numerical tools, *Markov Chain Monte Carlo* (MCMC) methods, which allow to simulate from the posterior distribution. We refer the reader to textbooks on Bayesian statistics, such as Robert (1994), Robert and Casella (1999), Casella and Berger (2001) or Frühwirth-Schnatter (2006). For the underlying paper it is sufficient to know that these econometric tools allow to sample from the posterior $\pi(\psi|x^t)$ whenever the likelihood $\pi(x^t|\psi)$ is available. The samples are denoted by $\psi^{[m]}$, $m = B+1, \ldots, M$. $B$ is the number of burn in steps which are required to attain convergence of the corresponding Markov chain. Then the conditional expectation $\mathbb{E}\left(\psi|.\right)$ will be derived by taking the sample mean of $\psi^{[m]}$.

Since we have to deal with a missing data problem in Section 3, we briefly remark on data augmentation, proposed in Tanner and Wong (1987). The trick of this method is quite simple: Suppose that $f(x|\psi)$ is not available in closed form but $f(x|\psi, y)$, where $y$ is a latent variable, a missing value or the realization of a latent process. By means of the Bayes theorem the posterior $\pi(\psi, y|x)$ is proportional to $f(x|\psi, y)\pi(\psi, y)$. By the above arguments, we are able to sample from the joint posterior $\pi(\psi, y|x)$, such that posterior samples of $\psi, y$ are available as well. This trick can be applied to a regression setting with missing values.

Using the notation of Section 2, consider the model $\Delta^\zeta z_t := z_t - \zeta z_{t-1} = (1-\zeta)\theta + \eta_t$, $\zeta \neq 1$ and $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, where the data $\Delta^\zeta z^t$ includes missing values. The regular observations are abbreviated by $\Delta^\zeta z^*$, the missing values are $\Delta^\zeta z^t \setminus \Delta^\zeta z^*$. Now augment the parameter space by $\Delta^\zeta z^{**}$, where each elements of $\Delta^\zeta z^{**}$ replaces the a missing observation of $\Delta^\zeta z^t$. Then the joint posterior of $\theta$ and $\tilde{z}^{**}$ can be derived by running a Bayesian sampler consisting of the following steps: For each sweep $m$:

`Step 1`: sample $\theta^{[m]}$ from $p(\tilde{\theta}|\Delta^\zeta z^*, \Delta^\zeta z^{**,[m-1]})$.

`Step 2`: sample $\Delta^\zeta z^{**,[m]}$ from $p(\Delta^\zeta z^{**}|\theta^{[m]}, \Delta^\zeta z^*)$.

Sampling `Step 1` exactly corresponds to drawing from the normal distribution described by (10). For the augmented missing values, $\Delta^\zeta z^{**}$, one can work with the Metropolis Hastings algorithm. Given $\theta$, $\Delta^\zeta z_t^{**}$ is normally distributed with mean $(1-\zeta)\theta$ and variance $\sigma_\eta^2$. the innovations $\eta_t$ are obtained recursively. By means of $\tilde{\theta}^{[m]}$ and $\Delta^\zeta z_t$ one directly obtains the residuals. The likelihood $f((\Delta^\zeta z^*, \Delta^\zeta z^{**})|\tilde{\theta})$ is given by the product of normals with mean $(1-\zeta)\theta$ and variance $\sigma_\eta^2$. Thus, one proposes $\Delta^\zeta z_t^{**,p}$ from a normal density $q(\Delta^\zeta z_t^{**,\cdot})$ with mean $(1-\zeta)\theta^{[m]}$ and variance $\sigma_\eta^2$ and accepts the proposed $\Delta^\zeta z_t^{**,p}$ with a probability of $min\{1, \frac{f((\Delta^\zeta z^*, \Delta^\zeta z^{**,p})|\tilde{\theta}^{[m]})}{f((\Delta^\zeta z^*, \Delta^\zeta z^{**,[m-1]})|\tilde{\theta}^{[m]})} \frac{q(\Delta^\zeta z^{**,[m-1]})}{q(\Delta^\zeta z^{**,p})}\}$, otherwise $\Delta^\zeta z^{**,[m]} = \Delta^\zeta z^{**,[m-1]}$. Here $\Delta^\zeta z^{**,[m-1]}$ are the current augmented missing values in the sampler. Alternative proposal densities can also be applied as well as forward-filtering backward sampling (see Frühwirth-Schnatter (1994)). If necessary, the missing $z_t$ can now be derived from $z_0$ and $\Delta^\zeta z_t$.

*Prediction and Missing Values*: If at time $t$, $z_{t-j}, \ldots, z_{t-1}$ are missing, then firms can take use of (14). By strictly sticking to Bayesian sampling $\Delta^\zeta z_{t-j-1}^{**}, \ldots, \Delta^\zeta z_{t-1}^{**}$ will be simulated by means of `Step 2`; missing $z_t^{**}$ follow from $\Delta^\zeta z_t$. By the structure of the model $z_{t+1}|z_t, \theta^{[m]}$ is normally distributed with mean $(1-\zeta)\theta^{[m]} + \zeta z_t$ and variance $\sigma_\eta^2$. $\theta^{[m]}$ is by (10) normally distributed with mean $a_{t-1}$ and variance $A_{t-1}$.

# References


Adam, K. and Marcet, A. (2010). Booms and busts in asset prices. IMES Discussion Paper Series 10-E-02, Institute for Monetary and Economic Studies, Bank of Japan.

Alfarano, S. and Milakovic, M. (2009). Network structure and n-dependence in agent-based herding models. *Journal of Economic Dynamics and Control*, 33(1):78 – 92.

Barron, A., Schervish, M. J., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Annals of Statistics*, 27(2):536–561.

Bickel, P. and Doksum, K. (2001). *Mathematical Statistics*. Prentice Hall, New Jersey, 2 edition.

Billingsley, P. (1986). *Probability and Measure*. Wiley, Wiley series in probability and mathematical statistics, New York, 2 edition.

Blume, L. and Easley, D. (1993). Rational expectations and rational learning. Game theory and information, EconWPA.

Blume, L. E. and Easley, D. (1998). Rational expectations and rational learning. In Mukul Majumdar, editor, Organizations with Incomplete Information, Cambridge University Press.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, 65(5):1059–1096.

Casella, G. and Berger, R. (2001). *Statistical Inference*. Duxbury Resource Center, Boston, 2 edition.

Chernozhukov, V. and Hong, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica*, 72(5):1445–1480.

Chib, S. (1993). Bayes regression with autoregressive errors. *Journal of Econometrics*, 58:77–99.

Christensen, R. (2009). Inconsistent bayesian estimation. *Bayesian Analysis*, 4:413–416.

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Annals of Statistics*, 14(1):1–26.

Evans, G. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton.

Feldman, M. (1991). On the generic nonconvergence of bayesian actions and beliefs. *Economic Theory*, 1:301–321.

Fernandez, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite dimensional parameters. *Technical Report, number 492, Department of Statistics, University of California Berkeley*.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Ghosal, S., , and Samanta, T. (1995a). Stability and convergence of posterior in non-regular problems. *Mathematical Methods of Statistics*, 4(4):361–388.

Ghosal, S. (2000). A review of consistency and convergence of posterior distribution. Technical report, In Proceedings of National Conference in Bayesian Analysis, Benaras Hindu University, Varanashi, India.

Ghosal, S., Ghosh, J. K., and Samanta, T. (1994). *Stability and Convergence of Posterior in Non-regular Problems*. Statistical Decision Theory and Related Topics 5, (S.S. Gupta and J.O. Berger, eds.), Springer, New York.

Ghosal, S., Ghosh, J. K., and Samanta, T. (1995b). On convergence of posterior distributions. *Annals of Statistics*, 23(6):2145–2152.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 28:500–531.

Ghosal, S. and Tang, Y. (2006). Bayesian consistency for markov processes. *Sankhya*, 68:227–239.

Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for non-iid observations. *Annals of Statistics*, 35:192–223.

Guesnerie, R. (1992). An exploration of the eductive justifications of the rational-expectations hypothesis. *American Economics Review*, 82(5):1254–1278.

Guesnerie, R. (1993). Theoretical tests of the rational expectations hypothesis in economic dynamical models. *Journal of Economic Dynamics and Control*, 17(5-6):847 – 864.

Guesnerie, R. and Jara-Moroni, P. (2009). Expectational coordination in simple economic contexts: Concepts and analysis with emphasis on strategic substitutabilities. Working Paper 2009 - 27, Paris School of Economics.

Harrison, J. and West, M. (1997). *Bayesian Forecasting and Dynamic Models.* Springer.

Hirano, K. and Porter, J. R. (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica*, 71(5):1307–1338.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical estimation, asymptotic theory.* Applications of Mathematics, vol. 16, Springer, New York.

Jang, G. H., Lee, J., and Lee, S. (2010). Posterior consistency of species sampling priors. *Statistica Sinica*, 20:581–593.

Jun, B. and Vives, X. (1996). Learning and convergence to a full-information equilibrium are not equivalent. *Review of Economic Studies*, 63(4):653–674.

Kalai, E. and Lehrer, E. (1992). Bayesian forecasting. Discussion paper 998, Department of Managerial Economics and Decision Sciences Kellog School of Management, Northwestern University.

Kelly, D. L. and Kolstad, C. D. (1999). Bayesian learning, growth, and pollution. *Journal of Economic Dynamics and Control*, 23(4):491 – 518.

Klenke, A. (2008). *Probability Theory - A Comprehensive Course.* Springer.

LeCam, L. M. and Yang, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York.

Lehmann, E. (1991). *Theory of Point Estimation*. Wadsworth and Brooks/Cole.

Lijoi, A., Prünster, I., and Walker, S. G. (2007). Bayesian consistency for stationary models. *Econometric Theory*, 23(4):749–759.

Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press (Cambridge Mathematical Library), New York, 2 edition.

Paolella, M. (2007). *Intermediate Probability - A Computational Approach*. Wiley.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models, Asymptotic Theory*. Springer, New York.

Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York.

Robert, C. P. (1994). *The Bayesian Choice*. Springer, New York.

Rodriguez-Yam, G., Davis, R. A., and Scharf, L. (2004). Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. Technical report, Department of Statistics, Columbia University.

Rousseau, J., Chopin, N., and Liseo, B. (2010). Bayesian nonparametric estimation of the spectral density of a long or intermediate memory gaussian process. Working paper, Université Paris Dauphine, Paris.

Routledge, B. (1999). Adaptive learning in financial markets. *The Review of Financial Studies*, 12:1165–1202.

Schwartz, L. (1965). On bayes procedures. *Probability Theory and Related Fields*, 4:10–26.

Shalizi, C. R. (2010). Frequentist consistency of bayesian procedures. Technical report, Notebook, University of Michigan, Department of Statistics, http://cscs.umich.edu/ crshalizi/notebooks/bayesian-consistency.html.

Smith, L. and Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398.

Stinchcombe, M. B. (2005). The unbearable flightiness of bayesians: Generaically erratic updating. Working paper, University of Texas at Austin.

Strasser, H. (1985). *Mathematical Theory of Statistics*. Studies in Mathematics 7. de Gruyter, Berlin.

Sun, Y. (2006). The exact law of large numbers via fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126(1):31 – 69.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–550.

Timmermann, A. (1996). Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *Review of Economic Studies*, 63(4):523575.

Vives, X. (2008). *Information and Learning in Markets*. Princeton University Press.

Walker, S. (2004). New approaches to bayesian consistency. *Annals of Statistics*, 32:2028–2043.