A Perturbation Approach to Nonlinear Filtering: The Case of Stochastic Volatility

Ivana Komunjer[*] Natalia Sizova[†]

This Draft: December 15, 2016

Abstract

Nonlinear filtering, that is the computation of the conditional distribution of a latent state vector given the available information, is generally an infinite dimensional problem for which no closed form solutions exist. This paper uses the perturbation method to derive an approximate filter for models in which the hidden state follows a linear Gaussian transition equation, but the observation equation is nonlinear and non-Gaussian. The proposed approximate filter is characterized by a finite number of sufficient statistics that are Markovian, and whose number increases with the order of the approximation. Explicit formulas are derived for stochastic volatility models.

Keywords: nonlinear state space systems; Bayes updating; stochastic volatility.

JEL Classification: C22; C58.

1 Introduction

This paper proposes a perturbation method to solve the nonlinear filtering problem, that is the problem of computing the conditional distribution of a latent state vector given the available information. While the traditional expansion based methods such as the extended Kalman filter (EKF) aim at linearizing the nonlinear observation equation (see, e.g., Anderson and Moore, 1979), our method approximates directly the object of interest: the conditional distribution of a latent state vector given the current information, also known in the literature as the *nonlinear filter*. The key features of the proposed method are: (i) closed form solutions for increasing orders of approximation, unlike in the case of EKF which is a first-order approximation method; (ii) no curse of dimensionality unlike in the case of discretization

[*]University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. Email: komunjer@ucsd.edu

[†]Rice University, 6100 Main Street, Houston, TX 77251. Email: natalia.sizova@rice.edu

Acknowledgments: This paper was presented at 2015 NBER-NSF Time Series conference in Vienna, 2015 Midwest Econometrics Group meeting at FRB of St. Louis, 2016 CEF conference in Bordeaux, as well as Econometrics seminars at Indiana University, University of Washington, USC, and Northwestern University. We thank seminar participants for helpful comments and suggestions.

or randomization methods such as particle filtering; (iii) ability to deliver reasonable approximations even in nonlinear state-space models that are not separable in the disturbances, case in which EKF fails to update based on observations. This feature is particularly relevant in the context of stochastic volatility models which are nonseparable in the observation errors. Finally, it is worth emphasizing that our approximation has a clear asymptotic interpretation. This feature is particularly important if one is interested in embedding our filtering method within an estimation procedure, or else within a structural dynamic stochastic choice model with imperfect information.

In the most general form, the problem of computing the conditional distribution of a latent state vector given the available information is an infinite dimensional problem for which no closed form solutions exist. Thus, some approximation to the nonlinear filter is needed. This problem has spun a vast literature that transgresses the boundaries of several fields. We limit our short review of this literature to the techniques applied in economics and finance. We start with the results that obtain in two special cases often assumed in structural economic models: discrete and linear Gaussian state space models. In their robust control model, Hansen and Sargent (2007, 2010) obtain closed-form solutions in the linear case and in the case with a discrete unobserved state with two values. Kalman filter can be used in the linear model. The optimal rule in the discrete model follows a simple Bayes updating that tracks two probabilities.

Stochastic volatility models, which are the leading example in this paper, are by construction nonlinear. Some approximation to the nonlinear filter in these models is therefore needed. Ruiz (1994) and Harvey, Ruiz, and Shephard (1994) apply the Kalman filter within the quasi-maximum likelihood estimation. Although the Kalman filter is not optimal in their setup, the resulting pseudo-posterior density can yield consistent quasi-maximum likelihood estimates of the model parameters. This approach can be applied to the dynamics that allow for linearization in the latent state after a suitable transformation, but cannot be extended to more general observation equations. Similar to other methods that extend the notion of the Kalman filter to non-linear settings, this method lacks asymptotic interpretation beyond the first-order approximation.

Jacquier, Polson, and Rossi (1994) suggest a new Bayesian (MCMC) approach to filtering stochastic volatility. This approach approximates the posterior density by a random sample. Shephard and Pitt (1997) and Kim, Shephard, and Chib (1998) offer improved algorithms for the method. The advantage of the Monte Carlo methods is that they provide an accurate way to generate posterior densities of the unobserved volatilities. However, the de-facto state variables for this approximation include all the random draws from the posterior distribution, which makes any Monte Carlo method difficult to incorporate within a dynamic stochastic choice model. Another similar approach that yields accurate approximations to the posterior density at the expense of computing time and tractability is the particle filter, also known as the sequential Monte Carlo method (see, e.g., Shephard and Pitt, 1999; Polson, Stroud, and Müller, 2008, for applications of the particle filter to the stochastic volatility models). Similarly to the MCMC approach, the particle filter depends on a large number of state variables, which complicates its use. One should note that the particle filter and Bayesian approaches have been widely applied to the structural models in macroeconomics (see, e.g., Fernández-Villaverde and Rubio-Ramírez, 2007, for the first application of the particle filter in this context). Creal (2012) provides a survey of sequential Monte Carlo methods in economics and finance.

The approximation method proposed in this paper is based on the perturbation approach. The main idea is simple: if the problem is smooth enough and admits a closed form solution in a special case, then an approximate solution can be obtained by considering small deviations of the problem around this case. The idea, dating back to the seminal paper by Fleming (1971), has been used to obtain approximate solutions to stochastic control problems around solutions obtained in the deterministic steady state (see, Bensoussan, 1988; Judd, 1996; Schmitt-Grohé and Uribe, 2004; Fernández-Villaverde and Rubio-Ramírez, 2006; Justiniano and Primicieri, 2008; Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Uribe, 2011; Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez, 2015). Unlike this prior work, we apply the perturbation method to approximate the nonlinear filter. The resulting approximation is a posterior density that at a given level accuracy depends on a finite number of sufficient statistics that are Markovian. In particular, we approximate the nonlinear filter around an uninformative case in which the observation density carries no information about the latent state. Our approximations take a form similar to the Edgeworth and Gram-Charlier expansions (see, e.g., Sorenson and Stubberud, 1968; Kizner, 1969). However, unlike these expansions which rely on the polynomial approximations to the dynamic equations, our method directly approximates the density functions. Hence, our method is invariant with respect to variable transformations. Finally, the Gram-Charlier expansions cannot be interpreted as true asymptotic approximations, unlike our perturbation approximation.

The leading example used in this paper is the stochastic volatility model, in which the return r_t and its volatility $\exp \sigma_t$ follow:

$$\begin{aligned} r_{t+1} &= \rho r_t + (\exp \sigma_t)\epsilon_{t+1}, \\ \sigma_{t+1} &= (1-\lambda)\bar{\sigma} + \lambda \sigma_t + \eta \omega_{t+1}, \end{aligned} \tag{1}$$

and where ϵ_t is iid with a density p_ϵ on $\mathbb{R}$ satisfying $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$, ω_t is iid standard Gaussian, and ϵ_t and ω_t are independent. The above is a central model of asset returns in the financial literature (see, e.g., Hull and White, 1987; Ghysels, Harvey, and Renault, 1996; Shephard, 2004). It also appears in the general equilibrium asset pricing models that capture the effect of uncertainty in the financial markets (see, e.g., Bansal and Yaron, 2004; Bollerslev, Tauchen, and Zhou, 2009); as well as in the macroeconomic (DSGE) models that study the negative effect of volatility on various macroeconomic indicators (see, e.g., Justiniano and Primicieri, 2008; Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Uribe, 2011; Caldara, Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Yao, 2011). Though all our derivations are specific to the model in (1), our method easily generalizes to state space models with linear Gaussian transition equations but possibly nonlinear and non-Gaussian observation equations.

The remainder of the paper is organized as follows. Section 2 describes the setup. The perturbation method and its application to the nonlinear filter is presented in Section 3. Section 4 discusses the robustness of our results and possible extensions. A Monte Carlo experiment is presented in Section 5. The final section concludes.

As a matter of notation, for any real function $f : \mathbb{R} \to \mathbb{R}$, $[f(x)]_+ = \max(f(x), 0)$ denotes the positive part of f, while $[f(x)]_- = -\min(f(x), 0)$ denotes the negative part of f. In what follows, $L_1(\mathbb{R})$ is the collection of equivalence classes of measurable real functions $f : \mathbb{R} \to \mathbb{R}$ whose L_1 norm $\|f\|_1 = \int |f| d\lambda$ is finite. Here, the measure space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ with λ denoting the Lebesgue measure. $L_1^+(\mathbb{R})$ denotes the positive cone in $L_1(\mathbb{R})$, i.e. $L_1^+(\mathbb{R}) = \{f \in L_1(\mathbb{R}) : f > 0\}$. We also let $\mathbb{R}^* \equiv \mathbb{R}\backslash\{0\}$. To denote derivatives of functions, we interchangeably use the notation: $\partial f(x)/\partial x = f'(x) = f_x(x) = f^{(1)}(x) = f_{(1)}(x)$, depending on the context. The degree of approximation is denoted by a bracketed superscript, as in $f^{[1]}(x)$, for example.

2 Model and Assumptions

We focus on a class of nonlinear and non-Gaussian state space models that take the form:

$$\begin{aligned} y_{t+1} &= \exp(\eta x_t)\epsilon_{t+1} \\ x_{t+1} &= \lambda x_t + \omega_{t+1}, \end{aligned} \tag{2}$$

for every $t \geqslant 0$, where $x_0 \in \mathbb{R}$ is drawn from a distribution with density p_0, $\{\epsilon_s\}_{s\geqslant 1}$ and $\{\omega_s\}_{s\geqslant 1}$ are iid sequences of random variables drawn from distributions with densities p_ϵ (mean 0 and variance 1) and φ (standard Gaussian), respectively, and $\{\epsilon_s\}_{s\geqslant 1}$, $\{\omega_s\}_{s\geqslant 1}$, and x_0 are independent. While we restrict ω_t to be Gaussian, we allow quite general choices for the distribution of ϵ_t. In particular, ϵ_t is not restricted to be Gaussian, and is allowed to exhibit thick tails, property particularly important in financial applications. The stochastic volatility model (1) is a special case of the model (2) obtained by letting $y_{t+1} = (r_{t+1} - \rho r_t)\exp(-\bar{\sigma})$ and $x_{t+1} = (\sigma_{t+1} - \bar{\sigma})/\eta$.

The two parameters appearing in the model (2) are: $\eta \in \mathbb{R}$ and $\lambda \in \mathbb{R}$, where the latter is further restricted to lie strictly inside the unit interval. The parameter η will play a particularly important role in our analysis, which is why we shall keep explicit reference to η in the expressions to follow. We shall however drop the reference to λ, and will treat its value as being fixed.

While we assume $y_t \in \mathbb{R}$ to be observed, $x_t \in \mathbb{R}$ remains latent.[1] Thus, the model in (2) is a hidden Markov model (see, e.g., Meyn and Tweedie, 1993, for definitions). Because of its observation equation, the model is both non-linear and non-Gaussian, with the observation density $p(y \mid x, \eta)$ given by

$$p(y \mid x, \eta) = \exp(-\eta x) p_\epsilon (\exp(-\eta x) y) \,.$$

The transition density $q(x' \mid x)$ is however Gaussian,

$$q(x' \mid x) = \varphi(x' - \lambda x).$$

Let $i_t = (y_1, \ldots, y_t)$ ($i_t \in I_t$) denote the information at time t, and let $h_t = (x_0, x_1, y_1, \ldots, x_t, y_t)$ ($h_t \in H_t$) denote the time t history. The time t information is the portion of the time t history known

[1]Extensions to models with higher-dimensional states x_t are straightforward, though at the expense of a more cumbersome notation.

at time t. We shall consider the density p_0 of x_0 as an element in $L_1(\mathbb{R})$. The object of interest is the conditional density of x_t given time t information i_t, object known in the literature as the *nonlinear filter*.[2] Hereafter, we maintain the following assumptions.

Assumption 1. *$p_0 > 0$ on $\mathbb{R}$ and $p_\epsilon > 0$ on $\mathbb{R}$.*

Assumption 2. *$|\lambda| < 1$.*

Assumption 3. *$p_\epsilon \in \mathcal{C}^1(\mathbb{R})$, $\sup_{u\in\mathbb{R}} |up_\epsilon(u)| < \infty$, and $\sup_{u\in\mathbb{R}} |u^2 p_\epsilon'(u)| < \infty$.*

The role of Assumption 1 is twofold: first, it ensures that the observation density $p(y \mid x, \eta)$ is everywhere strictly positive. Second, it implies that the sequence of probability densities $\pi_t(\eta, p_0) \in L_1(\mathbb{R}^{2t+1})$ of histories h_t defined by:

$$\pi_t(\eta, p_0)[x_0, x_1, y_1, \ldots, x_t, y_t] = p_0(x_0) \prod_{s=1}^{t} p(y_s \mid x_{s-1}, \eta) q(x_s \mid x_{s-1}),$$

satisfies $\pi_t(\eta, p_0) > 0$ on H_t. As a consequence, the conditional density $p_t(\eta, p_0)$ of x_t given time t information is well defined and given by:

$$p_t(\eta, p_0)[x_t] = \frac{\int_{\mathbb{R}^t} \pi_t(\eta, p_0)(x_0, x_1, y_1, \ldots, x_t, y_t) dx_0 dx_1 \ldots dx_{t-1}}{\int_{\mathbb{R}^{t+1}} \pi_t(\eta, p_0)(x_0, x_1, y_1, \ldots, x_t, y_t) dx_0 dx_1 \ldots dx_{t-1} dx_t}, \tag{3}$$

and has the property that $p_t(\eta, p_0) > 0$ on $\mathbb{R}$ for every $i_t \in I_t$. Equation (3) defines a functional process $\{p_t(\eta, p_0)\}_{t\geqslant 0}$ where each $p_t(\eta, p_0)$ is an element in $L_1(\mathbb{R})$ satisfying $p_t(\eta, p_0) > 0$ on $\mathbb{R}$ for every $i_t \in I_t$ (i.e. p_t is a probability density), that is parameterized by $i_t \in I_t$. In the terminology of Bertsekas and Shreve (1978, Definition 7.12), $p_t(\eta, p_0)$ is a stochastic kernel. Equation (3) also makes it clear that each $p_t(\eta, p_0)$ depends on the value of η as well as the choice of the prior p_0. Thus different choices of priors result in different processes $\{p_t(\eta, p_0)\}_{t\geqslant 0}$.

Under Assumption 2, the Markov chain $\{x_t\}_{t\geqslant 0}$ is positive recurrent, geometrically ergodic and its stationary distribution is Gaussian with mean 0 and variance $\sigma^2 \equiv (1 - \lambda^2)^{-1}$ (see, e.g. Meyn and Tweedie, 1993). In what follows, we denote by $\overline{p}$ the corresponding stationary density,

$$\overline{p}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

It then follows that $\{(x_t, y_{t+1})'\}_{t\geqslant 0}$ is a positive recurrent Markov chain with stationary distribution that has a density $\overline{p}(x)p(y \mid x, \eta)$. Existence of a unique stationary density for $\{x_t\}_{t\geqslant 0}$ shall be important for our purposes. It will allow us to pick $\overline{p}$ as a prior density for x_0, choice which proves convenient for computation purposes. As we shall later show in Section 4, our approximation results are robust to the choice of prior.

Assumption 3 plays several roles. First, by requiring that the probability density p_ϵ of the observation error ϵ_t be continuously differentiable on $\mathbb{R}$, this assumption implies that the observation density

[2] Some authors use the term "filter" differently, as a filtering algorithm or a posteriori state estimate.

$p(y \mid x, \eta)$ is a smooth enough function of the parameter η to justify the use of local approximation methods, such as Taylor's expansions around a particular value $\eta = 0$. Second, the boundedness conditions in Assumption 3 allow to define dominating functions that are instrumental in applying the Lebesgue dominated convergence theorem in our proofs. Lastly, this assumption combined with additional conditions on p_ϵ is also used to establish the irrelevance of the choice of prior.

3 Perturbation Approach

The starting point of our approach is a recursive formulation of the nonlinear filter. It is indeed straightforward to show (see, e.g., Lemma 10.4 Bertsekas and Shreve, 1978), that starting with a prior $\overline{p} \in L_1(\mathbb{R})$, the conditional density $p_t(\eta, \overline{p})$ in (3) can recursively be calculated as:[3]

$$\begin{aligned} p_1(\eta, \overline{p}) &= \phi(\overline{p}, y_1, \eta) \\ p_{t+1}(\eta, \overline{p}) &= \phi\left(p_t(\eta, \overline{p}), y_{t+1}, \eta\right), \quad t \geqslant 1 \end{aligned} \tag{4}$$

where $\phi : L_1(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \to L_1(\mathbb{R})$, and

$$\phi(f, y, \eta)[x'] \equiv \frac{\int q(x' \mid x) p(y \mid x, \eta) f(x) dx}{\int p(y \mid x, \eta) f(x) dx}. \tag{5}$$

In general, the recursion (4) cannot be solved analytically. The starting point of the perturbation method is to look for the case in which a closed form solution to (4) is available, then approximate p_t in a neighborhood of that particular solution. The solution which we consider is the unconditional density of x_t, obtained in the limit case in which the observation density $p(y \mid x, \eta)$ is not informative about the latent state. This amounts to approximating the conditional density p_t of the state around the case $\eta = 0$, in which case $p(y_{t+1} \mid x_t, 0) = p_\epsilon(y_{t+1})$. In this uninformative case, the solution to the filtering equation (4) is known and equal to

$$p_t(0, \overline{p}) = \overline{p} \quad \text{for all } t \geqslant 0.$$

The idea then is to use Taylor-series expansions of ϕ around $\eta = 0$ and recursively construct the corresponding perturbed densities. For example, at the first order, starting with $p_0^{[1]}(\eta, \overline{p}) = \overline{p}$ construct

$$p_1^{[1]}(\eta, \overline{p}) = \phi(p_0^{[1]}(\eta, \overline{p}), y_1, 0) + \phi_\eta(p_0^{[1]}(\eta, \overline{p}), y_1, 0)\eta.$$

Notice that on the right-hand side, we have replaced the nonlinear function $\phi(p_0^{[1]}(\eta, \overline{p}), y_1, \cdot)$ with its first order Taylor approximation around $\eta = 0$: $\phi(p_0^{[1]}(\eta, \overline{p}), y_1, 0) + \phi_\eta(p_0^{[1]}(\eta, \overline{p}), y_1, 0)\eta$. We then proceed to show that $p_1^{[1]}(\eta, \overline{p}) \in L_1(\mathbb{R})$, and that

$$\|p_1(\eta, \overline{p}) - p_1^{[1]}(\eta, \overline{p})\|_1 = o(|\eta|) \quad a.s. \tag{6}$$

[3] This recursive formulation for $p_t(\eta, p_0)$ remains of course valid even if we start with a different prior $p_0 \neq \overline{p}$. Here, for the reasons of prior irrelevance, we focus on the case $p_0 = \overline{p}$.

so that we can call $p_1^{[1]}(\eta,\overline{p})$ a first order approximation to $p_1(\eta,\overline{p})$. The construction of higher order approximations follows along the same lines, with higher order Taylor approximations to ϕ. The above construction clearly requires that $\phi(f,y,\cdot)$ be a differentiable function, property which we shall formally establish below.

Though the property in (6) guarantees that the approximate density $p_1^{[1]}(\eta,\overline{p})$ is close to $p_1(\eta,\overline{p})$ in L_1 distance as η gets small, this by itself does not guarantee that the moments of $p_1^{[1]}(\eta,\overline{p})$ approximate those of $p_t(\eta,\overline{p})$. To guarantee the latter, we would like a property of the kind

$$\left|\int_{\mathbb{R}} x^k p_1(\eta,\overline{p})[x]dx - \int_{\mathbb{R}} x^k p_1^{[1]}(\eta)[x]dx\right| = o(|\eta|), \quad a.s. \quad \text{for every } k \geqslant 1.$$

Since the above is implied by $\|(\cdot)^k p_1(\eta,\overline{p})[\cdot] - (\cdot)^k p_1^{[1]}(\eta,\overline{p})[\cdot]\|_1 = o(|\eta|)$ a.s., we are conducted to consider the differentiability properties of the mapping $\Phi^k : L_1(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \to L_1(\mathbb{R})$ where

$$\Phi^k(f,y,\eta)[x] = x^k \phi(f,y,\eta)[x], \quad k \geqslant 0. \tag{7}$$

Of course, when $k=0$, Φ^k simply reduces to ϕ.

Given $f \in L_1^+(\mathbb{R})$ and $y \in \mathbb{R}^*$, we first establish the Fréchet differentiability of $\Phi^k(f,y,\cdot) : \mathbb{R} \to L_1(\mathbb{R})$ on $\mathbb{R}$, i.e. that for every $\eta \in \mathbb{R}$ there exists a continuous linear map $\Phi^k_\eta(f,y,\eta) : \mathbb{R} \to L_1(\mathbb{R})$, such that

$$\lim_{h\to 0} \frac{\|\Phi^k(f,y,\eta+h) - \Phi^k(f,y,\eta) - \Phi^k_\eta(f,y,\eta)h\|_1}{|h|} = 0.$$

The following result formally establishes differentiability and provides an expression for the derivative.[4]

Lemma 1. *Let Assumptions 1 to 3 hold, and take any $k \geqslant 0$. Then, for any $f \in L_1^+(\mathbb{R})$ such that $\int |x|^{k+1} f(x)dx < \infty$, and any $y \in \mathbb{R}^*$, $\Phi^k(f,y,\cdot)$ is Fréchet differentiable on $\mathbb{R}$, with derivative*

$$\Phi^k_\eta(f,y,\eta)[x'] = (x')^k \left\{ \frac{\int q(x' \mid x) p_\eta(y \mid x,\eta) f(x)dx}{\int p(y \mid x,\eta) f(x) dx} - \frac{\left[\int q(x' \mid x) p(y \mid x,\eta) f(x)dx\right]\left[\int p_\eta(y \mid x,\eta) f(x)dx\right]}{\left[\int p(y \mid x,\eta) f(x)dx\right]^2} \right\}.$$

The above property is crucial in showing that not only $p_1^{[1]}(\eta,\overline{p})$ (and recursively constructed $p_t^{[1]}(\eta,\overline{p})$, $t \geqslant 1$) approximates $p_1(\eta,\overline{p})$ (and $p_t(\eta,\overline{p})$, $t \geqslant 1$) in L_1 norm as η gets small, but also that the moments of the first approximate the moments of the second. The proof of Lemma 1 relies on the Lebesgue dominated convergence theorem, which requires an integrable dominating function. The latter is easy to construct provided the first argument f of $\Phi^k(f,y,\eta)$ has bounded moment of order $k+1$, in a sense that $\int |x|^{k+1} f(x)dx < \infty$. Since we intend to apply the result of Lemma 1 recursively to the densities $p_t(\eta,\overline{p})$ it remains to establish the behavior of their moments. This is done in the following lemma.

Lemma 2. *Let Assumptions 1 to 3 hold, and take any $k \geqslant 1$. Then, for any $\eta \in \mathbb{R}$ and any initial density p_0 such that*

$$\int_{\mathbb{R}} |x|^k p_0(x)dx < \infty,$$

[4] We use subscripts to denote the partial derivatives with respect to a variable, i.e. $p_\eta(y \mid x,\eta) \equiv \partial p(y \mid x,\eta)/\partial\eta$.

the sequence of conditional densities $\{p_t(\eta, p_0)\}_{t\geqslant 1}$ *satisfies*

$$\int_{\mathbb{R}} |x|^k p_t(\eta, p_0)[x]dx < \infty \quad a.s. \quad \textit{for every } t \geqslant 1.$$

Put in words, Lemma 2 says that when it comes to integrability, the properties of the prior p_0 transfer to those of the entire sequence of conditional densities $\{p_t(\eta, p_0)\}_{t\geqslant 1}$ initiated at the prior p_0. In particular, letting $p_0 = \overline{p}$, which is Gaussian and satisfies $\int |x|^k \overline{p}(x)dx < \infty$ for every $k \geqslant 1$, ensures that for every $t \geqslant 1$ and every $k \geqslant 1$, $\int |x|^k p_t(\eta, \overline{p})[x]dx < \infty$ a.s.. Thus, we will be able to apply Lemma 1 to establish the differentiability of $\Phi^k(p_t(\eta, \overline{p}), y_{t+1}, \cdot)$.

We are now able to state our main result, which establishes the first order approximation to $p_t(\eta, \overline{p})$.

Theorem 1 (**First order approximation**). *Let Assumptions 1 to 3 hold. Consider*

$$p_t^{[1]}(\eta, \overline{p})[x] \equiv \overline{p}(x) \left[1 + A_{1,t}\eta x\right],$$

where

$$A_{1,t} = \lambda \left[A_{1,t-1} - \psi_1(y_t)\right], \quad t \geqslant 1, \quad A_{1,0} = 0,$$

and $\psi_1(y) = 1 + yp'_\epsilon(y)/p_\epsilon(y)$*. Then* $p_t^1(\eta, \overline{p}) \in L_1(\mathbb{R})$*,* $\int |p_t^{[1]}(\eta, \overline{p})(x)|dx = 1$*, and* $\int |x^k p_t^{[1]}(\eta, \overline{p})[x]|dx < \infty$*, a.s. and for every* $k \geqslant 1$*. Moreover,* $p_t^{[1]}(\eta, \overline{p})$ *is a* first order approximation *to* $p_t(\eta, \overline{p})$ *in a sense that for every* $t \geqslant 1$*,*

$$\|p_t(\eta, \overline{p}) - p_t^{[1]}(\eta, \overline{p})\|_1 = o(|\eta|) \quad a.s.$$
$$\int \left| x^k \left(p_t(\eta, \overline{p})[x] - p_t^{[1]}(\eta, \overline{p})[x] \right) \right| dx = o(|\eta|) \quad a.s. \quad \textit{for every } k \geqslant 1.$$

There are three important features of the above result. First, the approximate density $p_t^{[1]}(\eta, \overline{p})$ is summarized by one sufficient statistic $A_{1,t}$ which has the Markov property. There are few cases in which the exact conditional densities $p_t(\eta, \overline{p})$ are known to satisfy this condition. One example is a model with a discrete unobserved state that takes a few values, such as a model indicator in Hansen and Sargent (2010). Another example is a linear Gaussian state space model considered in Hansen and Sargent (2007). In this model, the sufficient statistics are the mean and variance of a normal density, which are updated through the Kalman filtering equations. Kalman filter is however no longer applicable in nonlinear models such as the one considered here. Theorem 1 shows that the updating equation for the sufficient statistic $A_{1,t}$ now explicitly depends on the density p_ϵ of the observation error through the ψ_1 function. For instance, if p_ϵ is Gaussian, then $\psi_1(y) = 1 - y^2$, while if p_ϵ is Student-t density with ν degrees of freedom, $\psi_1(y) = 1 - \frac{(\nu+1)y^2}{\nu-2+y^2}$.

Second, the density proposed in Theorem 1 is a *local* approximation to the exact nonlinear filter.[5] This feature of our approach is similar to that of the EKF, which too is a local approximation method (see, e.g. Anderson and Moore, 1979, for details). There are however fundamental differences. EKF

[5]As noted by Kim, Kim, Schaumburg, and Sims (2008), the local accuracy of perturbation methods is often sufficient in econometric applications involving simulations and forecasts.

relies on the linear approximations to observation and transition equations (rather then densities). If ϵ has zero mean, then the linearization of $y_{t+1} = \exp(\eta x_t)\epsilon_{t+1}$ around the filtered $\widehat{x}_t$ and the zero shock is $y_{t+1} \approx \exp(\eta \widehat{x}_t)\epsilon_{t+1}$. That is, the linearized system in the extended Kalman filter is uninformative about x_t. As a result, no observed information is utilized in updating the conditional density of x_t, and the latter remains equal to the prior $\overline{p}$. In contrast to the EKF technique, the perturbation approach expands densities. Therefore, even if the observation equation is nonseparable in the error ϵ_t, case often ruled out by the EKF literature, the approximate conditional densities of x_t utilize the information in y_t.

Third, the conditional densities $p_t^{[1]}(\eta, \overline{p})$ not only approximate $p_t(\eta, \overline{p})$ to the first order in L_1 norm, but it also holds that all the moments under $p_t^{[1]}(\eta, \overline{p})$ approximate those under $p_t(\eta, \overline{p})$. This follows from the second result of Theorem 1 since

$$\left| \int x^k p_t(\eta, \overline{p})[x] dx - \int x^k p_t^{[1]}(\eta, \overline{p})[x] dx \right| \leqslant \int \left| x^k \left(p_t(\eta, \overline{p})[x] - p_t^{[1]}(\eta, \overline{p})[x] \right) \right| dx.$$

The proof of Theorem 1 exploits the recursive construction of the approximation $p_t^{[1]}(\eta, \overline{p})$, and is by induction. The details are given in Appendix.

An important property that the approximation in Theorem 1 shares with Edgeworth and Gram-Charlier expansions (see, e.g., Sorenson and Stubberud, 1968; Kizner, 1969) is that the posterior density is nearly Gaussian. Though both expansions take a form of a polynomial expansion around the unconditional normal distribution, in the case of our approximation this is not an a priori assumption but rather an outcome of the Taylor decomposition with respect to η around zero. There is another key difference in the two types of expansions: while the Edgeworth and Gram-Charlier expansions result from polynomial approximations to the dynamic equations, our method directly approximates the density functions. Hence, our method is invariant with respect to variable transformations.

4 Extensions

4.1 Non-linear change of variables

As noted by Judd (2002), the performance of local perturbation methods can be substantially improved by appropriate changes of variables. The idea is to improve the performance of the perturbation approximation at large values of η by considering perturbations with respect to some nonlinear transformation of η. Currently, the literature provides little guidance on the optimal choice of transformations. Our proposal is to try different transformation families and choose the one that yields the smallest filter errors.

Consider a function $S : \mathbb{R} \to \mathbb{R}$, $\eta \mapsto S(\eta)$, such that $S(0) = 0$ and consider approximations to the nonlinear filter in terms of the new perturbation parameter $\varsigma \equiv S(\eta)$. If S is continuously differentiable and $S'(0) \neq 0$, we can locally invert S in a neighborhood of zero, and express η in terms of the new perturbation parameter $\varsigma = S(\eta)$. This simple transformation leads to the following result.

Corollary 1. *Let the assumptions of Theorem 1 hold. Consider a transformation $S \in \mathcal{C}^2(\mathbb{R})$ such that*

$S(0) = 0$ *and* $S'(0) \neq 0$. *Let*

$$p_t^{[1]}(\varsigma, \overline{p})[x] \equiv \overline{p}(x)\left[1 + \frac{A_{1,t}}{S'(0)}\varsigma x\right],$$

where

$$A_{1,t} = \lambda\left[A_{1,t-1} - \psi_1(y_t)\right], \quad t \geqslant 1, \quad A_{1,0} = 0.$$

Then $p_t^{[1]}(\varsigma, \overline{p})$ *is a first order approximation to* $p_t(\eta, \overline{p})$ *in* $\varsigma = S(\eta)$, *in a sense that for every* $t \geqslant 1$,

$$\|p_t(\eta, \overline{p}) - p_t^{[1]}(\varsigma, \overline{p})\|_1 = o(|\varsigma|) \quad a.s.$$

$$\int \left|x^k\left(p_t(\eta, \overline{p})[x] - p_t^{[1]}(\varsigma, \overline{p})[x]\right)\right| dx = o(|\varsigma|) \quad a.s. \quad \text{for every } k \geqslant 1.$$

As an example, we consider a transformation that maps $\eta \in \mathbb{R}$ to a bounded set. We define a particular transformation for which the new perturbation parameter $\varsigma = S(\eta)$ measures the degree of informativeness of the observation y_{t+1} about the latent state x_t. In particular, say that informativeness is measured by the signal-to-noise ratio. The signal about x_t is proportional to η^2. Denote the average magnitude of the noise expressed in the same units by s^2, $0 < s < \infty$. Therefore, the informativeness of y_{t+1} about x_t is measured by $\eta^2/(\eta^2 + s^2)$, which we set to be equal to $[S(\eta)]^2$. To make this transformation injective, we can define $S(\eta)$ so that it inherits the sign of η. Therefore, $\varsigma = S(\eta) \in (-1, 1)$ in the presence of the noise. Note that $S(0) = 0$ and $S'(0) = s^{-1} \neq 0$.

4.2 Higher order approximations

Establishing higher order approximations to the nonlinear filter hinges on the existence of higher order derivatives of the mapping $\Phi^k(f, y, \cdot)$ in (7). This in turn requires stronger differentiability and boundedness requirements on the density p_ϵ of the observation error in (2). In what follows, let $n \in \mathbb{N}$ denote the desired level of approximation.

Assumption 3'. $p_\epsilon \in \mathcal{C}^n(\mathbb{R})$, $\sup_{u \in \mathbb{R}} |u^{j+1} p_\epsilon^{(j)}(u)| < \infty$ *for* $j = 0, \ldots, n$.

Note that the conditions in Assumption 3 are equivalent to those of the new Assumption 3' obtained in the special case $n = 1$, i.e., the approximation is to the first order. Under this stronger differentiability assumption, we now have the following result on the differentiability of $\Phi^k(f, y, \cdot)$.

Lemma 3. *Let Assumptions 1, 2 and 3' hold, and take any* $k \geqslant 0$. *Then, for any* $f \in L_1^+(\mathbb{R})$ *such that* $\int |x|^{k+n} f(x)dx < \infty$, *and any* $y \in \mathbb{R}^*$, $\Phi^k(f, y, \cdot)$ *is n-times Fréchet differentiable on* $\mathbb{R}$, *with nth order derivative* $\Phi^k_{(n)}(f, y, \cdot)$ *that can be computed recursively as:*

$$\Phi^k_{(n)}(f, y, \eta) = \frac{1}{\int p(y \mid x, \eta) f(x)dx}\Bigg((x')^k \int q(x' \mid x) p_{(n)}(y \mid x, \eta) f(x)dx \\ - n! \sum_{j=1}^{n} \frac{\int p_{(n+1-j)}(y \mid x, \eta) f(x)dx}{(n+1-j)!} \frac{\Phi^k_{(j-1)}(f, y, \eta)}{(j-1)!}\Bigg),$$

with

$$p_{(n)}(y \mid x, \eta) = (-x)^n \exp(-\eta x) \sum_{j=0}^{n} \frac{n!}{(n-j)!j!} \sum \frac{j!}{b_1! \dots b_j!} \frac{1}{1!^{b_1} \dots j!^{b_j}} p_\epsilon^{(r)} (\exp(-\eta x) y) \left[y \exp(-\eta x) \right]^r,$$

where the second sum ranges over all different solutions in nonnegative integers $(b_1, \dots, b_j)$ *of* $b_1 + 2b_2 + \dots + jb_j = j$ *and where* r *is defined as* $r = b_1 + \dots + b_j$.

Compared with Lemma 1 which establishes the first order differentiability of $\Phi^k(f, y, \cdot)$, the above result requires stronger integrability conditions: one now needs $|x|^{k+n}$ to be integrable with respect to the first argument f of $\Phi^k(f, y, \eta)$, where n denotes the desired order of differentiation.

Based on the result of Lemma 3, the construction of higher order approximations to the nonlinear filter then proceeds as follows. For any desired order $n \geqslant 1$, start at $p_0^{[n]}(\eta, \overline{p}) = \overline{p}$, and construct

$$p_1^{[n]}(\eta, \overline{p}) = \phi(p_0^{[n]}(\eta, \overline{p}), y_1, 0) + \frac{1}{1!} \phi_{(1)}(p_0^{[n]}(\eta, \overline{p}), y_1, 0) \eta + \dots + \frac{1}{n!} \phi_{(n)}(p_0^{[n]}(\eta, \overline{p}), y_1, 0) \eta^n.$$

On the right-hand side of the above equality, we have now replaced the nonlinear function $\phi(p_0^{[n]}(\eta, \overline{p}), y_1, \cdot)$ with its nth order Taylor approximation around $\eta = 0$. The nth order differentiability result of Lemma 3 immediately implies that

$$\|p_1(\eta, \overline{p}) - p_1^{[n]}(\eta, \overline{p})\|_1 = o(|\eta|^n) \quad a.s.$$

and

$$\int \left| x^k \left(p_1(\eta, \overline{p})[x] - p_1^{[n]}(\eta, \overline{p})[x] \right) \right| dx = o(|\eta|^n) \quad a.s. \quad \text{for every } k \geqslant 1,$$

so that we can call $p_1^{[n]}(\eta, \overline{p})$ an nth order approximation to $p_1(\eta, \overline{p})$. The expressions for $p_t^{[n]}(\eta, \overline{p})$ are obtained recursively following the steps analogous to the ones used in the proof of Theorem 1.

Specifically, at order $n = 2$ one obtains:

$$p_t^{[2]}(\eta, \overline{p})[x] = \overline{p}(x) \left[1 + A_{1,t} \eta x + \frac{1}{2} \left(A_{2,t} + A_{1,t}^2 \right) \eta^2 (x^2 - \sigma^2) \right], \quad t \geqslant 1 \tag{8}$$

where $\sigma^2 = (1 - \lambda^2)^{-1}$

$$A_{1,t} = \lambda \left[A_{1,t-1} - \psi_1(y_t) \right], \quad t \geqslant 1, \quad A_{1,0} = 0$$
$$A_{2,t} = \lambda^2 \left[A_{2,t-1} + \psi_2(y_t) \right], \quad t \geqslant 1, \quad A_{2,0} = 0,$$

and $\psi_1(y) = 1 + y p_\epsilon'(y)/p_\epsilon(y)$, $\psi_2(y) = y p_\epsilon'(y)/p_\epsilon(\zeta_t) + y^2 \left[p_\epsilon''(y)/p_\epsilon(y) - (p_\epsilon'(y)/p_\epsilon(y))^2 \right]$.[6]

As before, the dynamics of the sufficient statistics $A_{1,t}$ and $A_{2,t}$ depends on the density p_ϵ of the

[6]Detailed derivations can be found in the Appendix.

shock ϵ_t in the observation equation. For example, if ϵ_t is Gaussian, the sufficient statistics evolve as:

$$\begin{aligned} A_{1,t} &= \lambda\left[A_{1,t-1} + y_t^2 - 1\right], \\ A_{2,t} &= \lambda^2\left[A_{2,t-1} - 2y_t^2\right] \end{aligned}$$

If on the other hand ϵ_t is Student t-distributed with ν degrees of freedom, then:

$$\begin{aligned} A_{1,t} &= \lambda\left[A_{1,t-1} + \frac{(\nu+1)y_t^2}{\nu-2+y_t^2} - 1\right], \\ A_{2,t} &= \lambda^2\left[A_{2,t-1} - 2\frac{(\nu+1)y_t^2}{\nu-2+y_t^2} + \frac{(\nu+1)y_t^4}{\left(\nu-2+y_t^2\right)^2}\right]. \end{aligned}$$

4.3 Choice of prior density

Our perturbation approximation in Theorem 1 is computed under the assumption that x_0 is drawn from the stationary distribution with density $\overline{p}$. This of course is likely to not be the case, which raises the question of sensitivity of our results to departures from $p_0 = \overline{p}$. We shall now show that the choice of the prior density is generally irrelevant, as the nonlinear filter $p_t(\eta, p_0)$ "forgets" the prior p_0 exponentially fast. For this, we impose the following additional assumptions.

Assumption 4. *$\int_{\mathbb{R}} |\ln|u|| p_\epsilon(u)du < \infty$.*

Assumptions 3 and 4 ensure that the nonlinear filter is stable, in a sense that its long run behavior does not depend on the choice of the initial density p_0. Assumption 4 in particular holds if the density p_ϵ remains bounded above at zero, and if is first moment is finite. The following result formalizes the notion of filter stability.

Lemma 4. *Let Assumptions 1-4 hold. For any initial distribution p_0 such that for some $\gamma > 0$,*

$$\int_{\mathbb{R}} \exp(\gamma|x|)p_0(x)dx < \infty, \tag{9}$$

there exists a positive constant $c > 0$ such that we have

$$\limsup_{t\to\infty} t^{-1} \ln \|p_t(\eta, p_0) - p_t(\eta, \overline{p})\|_1 < -c \quad a.s. \tag{10}$$

Put in words, the filter "forgets" the prior density p_0 exponentially fast. This property is particularly useful for our purposes because it guarantees that we can approximate the conditional density $p_t(\eta, p_0)$ started from any p_0 by approximating $p_t(\eta, \overline{p})$, i.e. the filter started from the stationary distribution.

5 Monte Carlo Experiment

In this section we assess the accuracy of the proposed perturbation filter through a Monte Carlo experiment. To measure the performance of our filtering technique we check how well the filters obtained at various orders of approximation match the conditional moments of the latent state vector. We consider the first two conditional moments of x_t and denote them by $\mu^*_{1,t} = \mathbb{E}(x_t|i_t)$ and $v^*_t = \mathbb{V}(x_t|i_t)$, respectively. Specifically,

$$\mu^*_{1,t} = \int x p_t(\eta, \bar{p})[x]dx \quad \text{and} \quad v^*_t = \int \left(x - \mu^*_{1,t}\right)^2 p_t(\eta, \bar{p})[x]dx.$$

Suppose, $\widehat{\mu}_{1,t}$ is the conditional mean from a candidate filter, and $\widehat{v}_t$ and $\widehat{\mu}_{2,t}$ are the conditional central and non central second moments, respectively. For example, for the perturbation filter of order n,

$$\widehat{\mu}_{1,t} = \int x p^{[n]}_t(\eta, \bar{p})[x]dx, \quad \widehat{\mu}_{2,t} = \int x^2 p^{[n]}_t(\eta, \bar{p})[x]dx, \quad \text{and} \quad \widehat{v}_t = \int \left(x - \widehat{\mu}_{1,t}\right)^2 p^{[n]}_t(\eta, \bar{p})[x]dx.$$

We use the following notations for the corresponding unconditional moments:

$$\mu^*_1 = \int x\bar{p}(x)dx, \quad \mu^*_2 = \int x^2\bar{p}(x)dx, \quad \text{and} \quad v^* = \int \left(x - \mu^*_1\right)^2 \bar{p}(x)dx.$$

Then the moment-based measures of accuracy are defined as follows:

$$\begin{aligned} e^I &= \frac{\mathbb{E}(\widehat{\mu}_{1,t} - \mu^*_{1,t})^2}{\mathbb{E}(\mu^*_1 - \mu^*_{1,t})^2}, \\ e^{II} &= \frac{\mathbb{E}(\widehat{\mu}_{1,t} - x_t)^2}{\mathbb{E}(\mu^*_1 - x_t)^2}, \\ e^{III} &= \frac{\mathbb{E}(\widehat{v}_t - v^*_t)^2}{\mathbb{E}(v^* - v^*_t)^2}, \\ e^{IV} &= \frac{\mathbb{E}(\widehat{\mu}_{2,t} - x^2_t)^2}{\mathbb{E}(\mu^*_2 - x^2_t)^2}. \end{aligned}$$

The error e^I measures how well the candidate filter approximates the first conditional moment of x_t relative to the uninformative case. It can be constructed for the observed data if we have a method to calculate the true conditional mean $\mu^*_{1,t}$. Alternatively, we can use the true realization of x_t, which is known in simulations, to construct the error e^{II}. This error is minimized at the true conditional mean, and, therefore, also measures the accuracy of estimating the first conditional moment. Further, we consider the error e^{III} that measures the accuracy of estimating the second moment relative to the uninformative case. Similarly to e^I, this measure relies on the knowledge of the true conditional moment v^*_t. In contrast, error e^{IV} depends on x^2_t which is observed in simulations. Note that according to the results of Theorem 1, each of the above errors should decrease with the order of approximation.

As a benchmark, we consider the performance of a filter with global convergence properties, namely the particle filter, also known as the sequential MCMC method. The particle filter is a global approximation method, in which the conditional density of x_t given i_t is discretized into a number of particles M. As $M \to \infty$, the conditional moments calculated from the discretized density converge to the true ones (see, e.g., Crisan and Doucet, 2002). We use the version of the particle filter with re-sampling that results in a stable filter (see, e.g., Douc, Fort, Moulines, and Priouret, 2009). In particular, the algorithm of the particle filter is as follows. We start the sequence with M random draws from the stationary distribution with density $\overline{p}$, $\{x_0^{(i)}\}_{i=1}^M$. The associated empirical probability function is denoted by $p^P(x_0|i_0)$. Recursively, for each period t, we update $p^P(x_t|i_t)$ to $p^P(x_t|i_{t+1})$ by re-sampling from $\{x_t^{(i)}\}_{i=1}^M$ using the importance sampling step. That is, we re-sample from $\{x_t^{(i)}\}_{i=1}^M$ with probability weights proportional to $p(y_{t+1}|x_t^{(i)}, \eta)$. Next, we move to the next period by sampling $x_{t+1}^{(i)}$ from $q(x_{t+1}|x_t^{(i)})$. The corresponding empirical probability function is denoted by $p^P(x_{t+1}|i_{t+1})$.

Note that, in our non-linear example, the actual conditional moments of x_t are not available in closed form. Therefore, the true values of $\mu_{1,t}^*$ and v_t^* are not known. However, they can be well approximated by the particle filter obtained using a large value of M. In simulations, we use $M = 10^5$ to estimate the true $\mu_{1,t}^*$ and v_t^*. Using $M = 10^4$ yields very close estimates of the conditional moments.

Another important metric that can be used to compare different filters is their tractability. In the absence of restrictions on tractability, the particle filter with $M = 10^5$ should be certainly preferred to all of the other methods compared in this section. However, the intended use of the proposed methods is within larger structural models in which agents make optimal decisions dynamically for each t. The particle filter with M particles results in a system with M additional Markov state variables, which complicates the model solution and interpretation. As a proxy for tractability we record the computational speed of different filtering methods, which we report relative to the speed of our perturbations methods, i.e., we normalize the average performance of our perturbation methods to 1. We compare computing times on a personal computer with Intel Core 2Quad processor 3.00GHz and 8.00GB of RAM and approximate the results to few digits.

For the perturbation filter we consider the first-, second-, and third-order perturbation filters in which $p_t(\eta, \bar{p})$ is approximated around the zero signal-to-noise ratio, as described in Section 4.1. The parameter that defines the relative level of the noise is calibrated in an independently simulated sample. For the particle filter, we consider a moderate number of particles $M = 10^5$ and a small number of particles $M = 10$. The latter depends on only 10 additional states and, therefore, it is a viable alternative to the perturbation filter in dynamic choice models with unobserved states.

For simulations we consider the stochastic volatility model in (1):

$$\begin{aligned} r_{t+1} &= \rho r_t + (\exp \sigma_t)\epsilon_{t+1}, \\ \sigma_{t+1} &= (1-\lambda)\bar{\sigma} + \lambda \sigma_t + \eta \omega_{t+1}, \end{aligned}$$

for which the latent state is $x_t = (\sigma_t - \bar{\sigma})/\eta$ and the observed value is $y_t = (r_t - \rho r_{t-1}) \exp(-\bar{\sigma})$. Under the first simulation scheme, ϵ_t is standard normal and, under the second simulation scheme, it is standardized t-distributed.

We calibrate parameters of the Gaussian model for the stock index data at three sampling frequencies. First, we use the estimates of the continuous-time model in Andersen, Benzoni, and Lund (2002) to obtain parameter values in daily units. This calibration is labeled as Model I. Second, we estimate the stochastic volatility model on monthly and annual data of the CRSP value-weighted index from January, 1926 to December, 2015 provided by the Wharton Research Data Services (WRDS). The parameters are obtained by likelihood maximization. On monthly data, the estimates are labeled as Model II and for annual data, the estimates are labeled as Model III. All of the parameters are summarized in Table 1.

Table 1: Monte Carlo Simulations Parameters

Parameters	Model I	Model II	Model III
ρ	0	0.03	-0.04
$\bar{\sigma}$	-0.316	-2.70	-1.70
λ	0.969	0.93	0.80
η	0.056	0.15	0.19

For the model with t-distributed shocks, ϵ_t is standardized $t(\nu)$ where $\nu = 1/0.139$ as estimated by Bollerslev (1987) for S&P 500 returns. We run 10000 simulations divided in four independent samples with 200 additional burn-in observations each. The comparison results are given in Tables 2 and 3.

It is clear that the particle filter with 10,000 particles approximates the true density very well with mean-squared error for the first conditional moment between 0.001 and 0.003. However, this method is on average 2×10^4 slower than perturbation methods. The particle filter with a small number of particles $M = 10$ runs faster but it is still 120 slower than the analytic methods and its accuracy in matching the conditional moments is worse than the accuracy of the unconditional mean and variance. For the model with normally distributed innovations, we need particle filter with at least $M = 100$ in order to match the performance of the perturbation filter; and with M around 200 - 500 in the case with t-distributed ϵ_{t+1}. The results with $M = 100$ are also provided in Tables 2 and 3.

The performance of the perturbation methods depends on the parameter values of the model and the distribution of the shocks. Notably, the mean-square error in matching the first-order moment by the third-order perturbation filter is between 0.01 and 0.09 for the t-distributed shocks, while it is between 0.169 and 0.324 for the normally-distributed shocks.

The performance of the perturbation methods as measured by errors e^{II} and e^{IV} is less sensitive to the distributional assumptions and closer to the performance of the particle filter with $M = 10,000$. Based on the metric e^{II} , the R^2 in predicting x_t by the particle filter with $M = 10,000$ is between 0.15 and 0.39. The same R^2 for the third-order approximation filter is between 0.14 and 0.35. Based on the metric e^{IV}, the R^2 in predicting x_t^2 by the particle filter with $M = 10,000$ is between 0.14 and 0.33. The same R^2 for the third-order approximation filter is between 0.13 and 0.29. Overall, the Monte Carlo evidence supports the theoretical results regarding the behavior of the proposed perturbation filter, and indicates that in the stochastic volatility models considered, third-order approximations perform equally well as a particle filter with a large number of particles $(10,000)$, however with substantially faster computation speeds (1:20,000).

6 Conclusion

This paper proposes a new local approximation method that can be used to approximate the nonlinear filter at increasing orders of accuracy. As for any local method, one expects the behavior of the approxi-

Table 2: Simulations: filters comparison, normal distribution

The table reports accuracy and computational speed of perturbation and particle filters in simulations from model (1) with normally distributed shock ϵ_{t+1} . Consider $\mu^*_{1,t}$ that is the estimate of $\mathbb{E}(x_t|i_t)$ based on a particle filter with $M = 10^5$ particles. Also, v^*_t is the estimate of $\mathbb{V}(x_t|i_t)$. For each method we estimate the corresponding $\widehat{\mu}_{1,t}$ and $\widehat{v}_t$. Additionally, we compute $\widehat{\mu}_{2,t}$ that estimates $\mathbb{E}(x^2_t|i_t)$. The approximation errors are defined as follows: $e^I = \mathbb{E}(\widehat{\mu}_{1,t} - \mu^*_{1,t})^2/\mathbb{E}(\mu^*_1 - \mu^*_{1,t})^2$, $e^{II} = \mathbb{E}(\widehat{\mu}_{1,t} - x_t)^2/\mathbb{E}(\mu^*_1 - x_t)^2$, $e^{III} = \mathbb{E}(\widehat{v}_t - v^*_t)^2/\mathbb{E}(v^* - v^*_t)^2$, and $e^{IV} = \mathbb{E}(\widehat{\mu}_{2,t} - x^2_t)^2/\mathbb{E}(\mu^*_2 - x^2_t)^2$. The parameters of the simulations are given in Table 1. The number of particles in the particle filter is denoted as M. The perturbation parameter for the perturbation filter is the signal-to-noise ratio. Running times are given relative to the average running time of the perturbation filter.

Method	Running Time	e^I	e^{II}	e^{III}	e^{IV}
Model I					
Particle M = 10,000	20000×	0.003	0.609	0.002	0.668
Particle M = 100	1500×	0.704	0.919	0.156	0.913
Particle M = 10	120×	4.453	2.465	1.011	1.457
Perturbation I-order	1	1.022	1.093	1	0.892
Perturbation II-order	1	1.022	1.093	0.267	0.857
Perturbation III-order	1	0.169	0.650	0.267	0.775
Model II					
Particle M = 10,000	20000×	0.002	0.614	0.002	0.622
Particle M = 100	1500×	0.236	0.712	0.111	0.719
Particle M = 10	120×	2.012	1.366	0.644	1.336
Perturbation I-order	1	0.835	0.940	1	0.818
Perturbation II-order	1	0.835	0.940	0.649	0.821
Perturbation III-order	1	0.324	0.738	0.649	0.725
Model III					
Particle M = 10,000	20000×	0.003	0.829	0.011	0.839
Particle M = 100	1500×	0.381	0.901	0.583	0.910
Particle M = 10	120×	2.194	1.213	3.589	1.222
Perturbation I-order	1	0.373	0.886	1	0.874
Perturbation II-order	1	0.373	0.886	0.499	0.875
Perturbation III-order	1	0.207	0.860	0.499	0.873

Table 3: Simulations: filters comparison, t-distribution

The table reports accuracy and computational speed of perturbation and particle filters in simulations from model (1) with t-distributed shock ϵ_{t+1} . Consider $\mu^*_{1,t}$ that is the estimate of $\mathbb{E}(x_t|i_t)$ based on a particle filter with $M = 10^5$ particles. Also, v^*_t is the estimate of $\mathbb{V}(x_t|i_t)$. For each method we estimate the corresponding $\widehat{\mu}_{1,t}$ and $\widehat{v}_t$. Additionally, we compute $\widehat{\mu}_{2,t}$ that estimates $\mathbb{E}(x^2_t|i_t)$. The approximation errors are defined as follows: $e^I = \mathbb{E}(\widehat{\mu}_{1,t} - \mu^*_{1,t})^2/\mathbb{E}(\mu^*_1 - \mu^*_{1,t})^2$, $e^{II} = \mathbb{E}(\widehat{\mu}_{1,t} - x_t)^2/\mathbb{E}(\mu^*_1 - x_t)^2$, $e^{III} = \mathbb{E}(\widehat{v}_t - v^*_t)^2/\mathbb{E}(v^* - v^*_t)^2$, and $e^{IV} = \mathbb{E}(\widehat{\mu}_{2,t} - x^2_t)^2/\mathbb{E}(\mu^*_2 - x^2_t)^2$. The parameters of the simulations are given in Table 1. The number of particles in the particle filter is denoted as M. The perturbation parameter for the perturbation filter is the signal-to-noise ratio. Running times are given relative to the average running time of the perturbation filter.

Method	Running Time	e^I	e^{II}	e^{III}	e^{IV}
Model I					
Particle M = 10,000	20000×	0.001	0.656	0.003	0.717
Particle M = 100	1500×	2.413	1.493	0.441	1.127
Particle M = 10	120×	4.884	1.953	4.391	3.876
Perturbation I-order	1	0.514	0.861	1	0.845
Perturbation II-order	1	0.514	0.861	0.082	0.819
Perturbation III-order	1	0.066	0.682	0.082	0.791
Model II					
Particle M = 10,000	20000×	0.002	0.656	0.002	0.662
Particle M = 100	1500×	0.415	0.779	0.196	0.785
Particle M = 10	120×	7.724	3.405	1.166	2.759
Perturbation I-order	1	0.616	0.890	1	0.833
Perturbation II-order	1	0.616	0.698	0.134	0.842
Perturbation III-order	1	0.090	0.698	0.134	0.712
Model III					
Particle M = 10,000	20000×	0.003	0.853	0.012	0.861
Particle M = 100	1500×	0.623	0.929	1.201	0.936
Particle M = 10	120×	8.528	2.172	8.243	1.935
Perturbation I-order	1	0.128	0.875	1	0.879
Perturbation II-order	1	0.128	0.875	0.090	0.880
Perturbation III-order	1	0.010	0.856	0.090	0.865

mation to decay for large deviations of the perturbation parameter. In order to extend the radius of good performance of the perturbation approximation, one or more of the following standard adjustments may be applied: Padé instead of Taylor approximations (e.g., Judd, 1996; Baker and Graves-Morris, 1996), change of variables (Judd, 2002; Fernández-Villaverde and Rubio-Ramírez, 2006), and using conditionally local approximations along the path of x_t. The latter is especially promising as it allows for larger deviations from the steady state.[7] The conditional perturbation approach starts with an assumption that the time t conditional density of x_t given the information i_t can substantially deviate from the unconditional density, but all the subsequent shocks are small. Note that this conditional approach is markedly different from the methods which linearize the dynamics around the filtered value of the state vector. The latter technique involves only changing the reference point from the unconditional expectation to the conditional expectation. In contrast, the truly conditional approach redefines the entire reference distribution (see, e.g., Sorenson and Stubberud, 1968, for a similar idea). We view this conditional approximation approach as an interesting extension of the work presented here, and we leave it for future research.

[7]We are grateful to Thomas J. Sargent for pointing this out.

A Proofs

Proof of Lemma 1. Consider the mapping $\Phi^k : L_1(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \to L_1(\mathbb{R})$ given by:

$$\Phi^k(f, y, \eta)[x'] \equiv (x')^k \frac{\int q(x' \mid x) p(y \mid x, \eta) f(x) dx}{\int p(y \mid x, \eta) f(x) dx}, \quad k \geqslant 0$$

with

$$\begin{aligned} p(y \mid x, \eta) &= \exp(-\eta x) p_\epsilon \left(\exp(-\eta x) y\right) \\ q(x' \mid x) &= \varphi(x' - \lambda x). \end{aligned}$$

Throughout the proofs we shall use the following implication of Assumption 3 and the property in (**??**): for any $y \in \mathbb{R}^*$,

$$p(y \mid x) = \frac{1}{|y|} \left| \exp(-\eta x) y p_\epsilon(\exp(-\eta x) y) \right|,$$

so for any $y \in \mathbb{R}^*$,

$$\sup_{x \in \mathbb{R}} p(y \mid x) \leqslant \frac{1}{|y|} \sup_{u \in \mathbb{R}} |u p_\epsilon(u)| < \infty.$$

We first need to show that Φ^k is well defined. For this, consider first the denominator: for any $f \in L_1^+(\mathbb{R})$, $\eta \in \mathbb{R}$ and $y \in \mathbb{R}^*$ we have

$$0 < \int p(y \mid x, \eta) f(x) dx \leqslant \frac{\sup_{u \in \mathbb{R}} |u p_\epsilon(u)|}{|y|} \int f(x) dx < \infty.$$

Similarly, for the numerator

$$\left| (x')^k \int q(x' \mid x) p(y \mid x, \eta) f(x) dx \right| \leqslant |x'|^k \varphi(0) \frac{\sup_{u \in \mathbb{R}} |u p_\epsilon(u)|}{|y|} \int f(x) dx < \infty.$$

Moreover,

$$\int |\Phi^k(f, y, \eta)[x']| dx' = \frac{\int \left| \int (x')^k q(x' \mid x) p(y \mid x, \eta) f(x) dx \right| dx'}{\int p(y \mid x, \eta) f(x) dx},$$

so for $k = 0$,

$$\Phi^0(f, y, \eta) > 0 \quad \text{and} \quad \int \Phi^0(f, y, \eta)[x'] dx' = 1,$$

so $\Phi^0(f, y, \eta) \in L_1^+(\mathbb{R})$. For $k \geqslant 1$, using the fact that for $a, b \in \mathbb{R}$, $(|a| + |b|)^k \leqslant C_k \left(|a|^k + |b|^k \right)$ for some $1 \leqslant C_k < \infty$, we have

$$
\begin{aligned}
\int |\Phi^k(f,y,\eta)[x']|dx' &\leqslant \frac{\int\int |x'|^k q(x' \mid x)p(y \mid x,\eta)f(x)dxdx'}{\int p(y \mid x,\eta)f(x)dx} \\
&\leqslant \frac{\int\int \left(|x'-\lambda x| + |\lambda x|\right)^k q(x' \mid x)p(y \mid x,\eta)f(x)dxdx'}{\int p(y \mid x,\eta)f(x)dx} \\
&\leqslant C_k \left\{ \frac{\int\int |x'-\lambda x|^k q(x' \mid x)p(y \mid x,\eta)f(x)dxdx'}{\int p(y \mid x,\eta)f(x)dx} + \frac{\int |\lambda x|^k p(y \mid x,\eta)f(x)dx}{\int p(y \mid x,\eta)f(x)dx} \right\} \\
&\leqslant C_k \left\{ \int |u|^k \varphi(u)du + \frac{\frac{1}{|y|}\sup_{u\in\mathbb{R}} |up_\epsilon(u)|}{\int p(y \mid x,\eta)f(x)dx} \int |\lambda x|^k f(x)dx \right\} \\
&\leqslant C_k \left\{ \frac{2^{k/2}}{\sqrt{\pi}}\Gamma\left(\frac{k+1}{2}\right) + \frac{\frac{1}{|y|}\sup_{u\in\mathbb{R}} |up_\epsilon(u)|}{\int p(y \mid x,\eta)f(x)dx} \int |\lambda x|^k f(x)dx \right\} \\
&< \infty \quad a.s.
\end{aligned}
$$

for any $f \in L_1^+(\mathbb{R})$ such that $\int |x|^k f(x)dx < \infty$. This establishes that $\Phi^k(f,y,\eta) \in L_1(\mathbb{R})$ for $k \geqslant 1$.

Now, fix any $f \in L_1^+(\mathbb{R})$ and $y \in \mathbb{R}^*$, and write

$$
\Phi^k(f,y,\eta)[x] = \frac{N^k(f,y,\eta)[x]}{D(f,y,\eta)}
$$

with $N^k(f,y,\cdot) : \mathbb{R} \to L_1(\mathbb{R})$ and $D(f,y,\cdot) : \mathbb{R} \to \mathbb{R}$ given by

$$
N^k(f,y,\eta)[x'] \equiv (x')^k \int q(x' \mid x)p(y \mid x,\eta)f(x)dx, \quad D(f,y,\eta) \equiv \int p(y \mid x,\eta)f(x)dx.
$$

Note that $D(f,y,\cdot)$ is a real function such that $D(f,y,\cdot) > 0$ on $\mathbb{R}$, so we can use the product rule for Fréchet derivatives to show that if $N^k(f,y,\cdot) : \mathbb{R} \to L_1(\mathbb{R})$ and $D(f,y,\cdot) : \mathbb{R} \to \mathbb{R}$ are differentiable on $\mathbb{R}$ with derivatives $N_\eta^k(f,y,\cdot) : \mathbb{R} \to L_1(\mathbb{R})$ and $D_\eta(f,y,\cdot) : \mathbb{R} \to \mathbb{R}$, respectively, so is $\Phi^k(f,y,\cdot)$ and

$$
\Phi_\eta^k(f,y,\eta) = \frac{N_\eta^k(f,y,\eta)}{D(f,y,\eta)} - \frac{N^k(f,y,\eta)D_\eta(f,y,\eta)}{[D(f,y,\eta)]^2}
$$

We now show that both $N^k(f,y,\cdot)$ and $D(f,y,\cdot)$ are differentiable on $\mathbb{R}$. We start with $D(f,y,\cdot)$. Since $p_\epsilon \in \mathcal{C}^1(\mathbb{R})$, we have

$$
\begin{aligned}
p_\eta(y \mid x,\eta) &\equiv \frac{\partial p(y \mid x,\eta)}{\partial \eta} \\
&= -x\exp(-\eta x)\left[p_\epsilon\left(\exp(-\eta x)y\right) + \exp(-\eta x)yp'_\epsilon\left(\exp(-\eta x)y\right)\right]
\end{aligned}
$$

so that for every $y \in \mathbb{R}^*$,

$$
|p_\eta(y \mid x,\eta)| \leqslant \frac{|x|}{|y|}\left[\sup_{u\in\mathbb{R}} |up_\epsilon(u)| + \sup_{u\in\mathbb{R}} |u^2 p'_\epsilon(u)|\right]
$$

So if

$$\int_{\mathbb{R}} |x| f(x) dx < \infty,$$

Lebesgue dominated convergence theorem implies that $D(f, y, \cdot)$ is differentiable on $\mathbb{R}$ with

$$D_\eta(f, y, \eta) = \int p_\eta(y \mid x, \eta) f(x) dx. \tag{11}$$

To establish the differentiability of $N^k(f, y, \cdot)$, we consider the limit as $h \to 0$ of

$$\begin{aligned}
&\frac{1}{|h|} \left\| \int (\cdot)^k q(\cdot \mid x) p(y \mid x, \eta + h) f(x) dx - \int (\cdot)^k q(\cdot \mid x) p(y \mid x, \eta) f(x) dx - h \int (\cdot)^k q(\cdot \mid x) p_\eta(y \mid x, \eta) f(x) dx \right\|_1 \\
&= \frac{1}{|h|} \left\| \int (\cdot)^k q(\cdot \mid x) \left[p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta) \right] f(x) dx \right\|_1 \\
&= \frac{1}{|h|} \int \left| (x')^k \int q(x' \mid x) \big[p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta) \big] f(x) dx \right| dx' \\
&\leqslant \int \int |x'|^k q(x' \mid x) \frac{\big| p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta) \big|}{|h|} f(x) dx dx'.
\end{aligned} \tag{12}$$

Consider first the value $k = 0$; then from (12) we have

$$\begin{aligned}
&\frac{1}{|h|} \left\| \int q(\cdot \mid x) p(y \mid x, \eta + h) f(x) dx - \int q(\cdot \mid x) p(y \mid x, \eta) f(x) dx - h \int q(\cdot \mid x) p_\eta(y \mid x, \eta) f(x) dx \right\|_1 \\
&\leqslant \int \int q(x' \mid x) \frac{\big| p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta) \big|}{|h|} f(x) dx dx' \\
&= \int \frac{\big| p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta) \big|}{|h|} f(x) dx \\
&= \int \big| p_\eta(y \mid x, \eta + h^*) - p_\eta(y \mid x, \eta) \big| f(x) dx,
\end{aligned}$$

since by the mean value theorem there exists $h^* \in (0, h)$ such that

$$p(y \mid x, \eta + h) - p(y \mid x, \eta) = h p_\eta(y \mid x, \eta + h^*).$$

Now,

$$\begin{aligned}
\big| p_\eta(y \mid x, \eta + h^*) - p_\eta(y \mid x, \eta) \big| &\leqslant \big| p_\eta(y \mid x, \eta + h^*) \big| + \big| p_\eta(y \mid x, \eta) \big| \\
&\leqslant 2 \frac{|x|}{|y|} \left[\sup_{u \in \mathbb{R}} |u p_\epsilon(u)| + \sup_{u \in \mathbb{R}} |u^2 p'_\epsilon(u)| \right]
\end{aligned}$$

so if

$$\int_{\mathbb{R}} |x| f(x) dx < \infty,$$

we can again use the Lebesgue dominated convergence theorem to pass the limit inside the integral and establish that $N^0(f, y, \cdot)$ is differentiable on $\mathbb{R}$ with derivative

$$N_\eta^0(f, y, \eta)[x'] = \int q(x' \mid x) p_\eta(y \mid x, \eta) f(x) dx. \tag{13}$$

Now consider the case $k \geqslant 1$ in (12):

$$\begin{aligned}
&\frac{1}{|h|} \left\| \int (\cdot)^k q(\cdot \mid x) p(y \mid x, \eta + h) f(x) dx - \int (\cdot)^k q(\cdot \mid x) p(y \mid x, \eta) f(x) dx - h \int (\cdot)^k q(\cdot \mid x) p_\eta(y \mid x, \eta) f(x) dx \right\|_1 \\
&\leqslant \int \int \left(|x' - \lambda x| + |\lambda x|\right)^k q(x' \mid x) \frac{\left|p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta)\right|}{|h|} f(x) dx dx' \\
&\leqslant C_k \int \left[\frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) + |\lambda x|^k \right] \frac{\left|p(y \mid x, \eta + h) - p(y \mid x, \eta) - h p_\eta(y \mid x, \eta)\right|}{|h|} f(x) dx,
\end{aligned}$$

where as before C_k is a constant ($1 \leqslant C_k < \infty$) such that $(|a| + |b|)^k \leqslant C_k \left(|a|^k + |b|^k\right)$ with $a, b \in \mathbb{R}$. Then, using a reasoning similar to that above, if both $\int |x| f(x) dx < \infty$ and $\int |x|^{k+1} f(x) dx < \infty$, it follows that $N^k(f, y, \cdot)$ is differentiable on $\mathbb{R}$ with derivative

$$N_\eta^k(f, y, \eta)[x'] = (x')^k \int q(x' \mid x) p_\eta(y \mid x, \eta) f(x) dx, \quad k \geqslant 1. \tag{14}$$

Combining both results (13) and (14) then gives

$$N_\eta^k(f, y, \eta)[x'] = (x')^k \int q(x' \mid x) p_\eta(y \mid x, \eta) f(x) dx, \quad k \geqslant 0. \tag{15}$$

Finally, combining (11) and (15) then establishes that for $k \geqslant 0$, $\Phi^k(f, y, \cdot)$ is differentiable on $\mathbb{R}$ with derivative:

$$\Phi_\eta^k(f, y, \eta)[x'] = (x')^k \left\{ \frac{\int q(x' \mid x) p_\eta(y \mid x, \eta) f(x) dx}{\int p(y \mid x, \eta) f(x) dx} - \frac{\left[\int q(x' \mid x) p(y \mid x, \eta) f(x) dx\right] \left[\int p_\eta(y \mid x, \eta) f(x) dx\right]}{\left[\int p(y \mid x, \eta) f(x) dx\right]^2} \right\}.$$

□

Proof of Lemma 2. The reasoning is by induction. Assume that for $t \geqslant 0$, $p_t(\eta, p_0)$ is such that

$$\int_{\mathbb{R}} |x|^k p_t(\eta, p_0)[x] dx < \infty \quad a.s. \tag{16}$$

We now proceed to show that $p_{t+1}(\eta, p_0)$ then also satisfies the above property. For this, note that

$$\int_{\mathbb{R}} |x'|^k p_{t+1}(\eta, p_0)[x'] dx' = \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} |x'|^k q(x' \mid x) p(y \mid x, \eta) p_t(\eta, p_0)[x] dx dx'}{\int_{\mathbb{R}} p(y \mid x, \eta) p_t(\eta, p_0)[x] dx}$$

Now, using the fact that for any $k \geqslant 1$ there exists $C_k \geqslant 1$ such that $\left(|a| + |b|\right)^k \leqslant C_k\left(|a|^k + |b|^k\right)$ (where

$a, b \in \mathbb{R}$), we have

$$\begin{aligned}\int_{\mathbb{R}} |x'|^k q(x' \mid x) dx' &\leqslant \int_{\mathbb{R}} [|x' - \lambda x| + |\lambda x|]^k q(x' \mid x) dx' \\ &\leqslant C_k \left[\int_{\mathbb{R}} |x' - \lambda x|^k q(x' \mid x) dx' + |\lambda x|^k \right] \\ &= \int_{\mathbb{R}} |u|^k \varphi(u) du + |\lambda x|^k \\ &= 2^{k/2}/\sqrt{\pi}\Gamma((k+1)/2) + |\lambda x|^k,\end{aligned}$$

so for any $y \in \mathbb{R}^*$,

$$\begin{aligned}\int_{\mathbb{R}} |x'|^k p_{t+1}(\eta, p_0)[x'] dx' &\leqslant \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) + \frac{\int_{\mathbb{R}} |\lambda x|^k p(y \mid x, \eta) p_t(\eta, p_0)[x] dx}{\int_{\mathbb{R}} p(y \mid x, \eta) p_t(\eta, p_0)[x] dx} \\ &\leqslant \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) + \frac{\frac{1}{|y|} \sup_{u \in \mathbb{R}} |u p_\epsilon(u)|}{\int_{\mathbb{R}} p(y \mid x, \eta) p_t(\eta, p_0)[x] dx} \int_{\mathbb{R}} |\lambda x|^k p_t(\eta, p_0)[x] dx \\ &< \infty,\end{aligned}$$

where the last inequality follows from Assumptions 2, 3 and from the property in (16). This then establishes the desired property at $t+1$. $\square$

Proof of Lemma 3. As in the proof of Lemma 1, we write

$$\Phi^k(f, y, \eta)[x] = \frac{N^k(f, y, \eta)[x]}{D(f, y, \eta)}$$

with $f \in L_1^+(\mathbb{R})$, $y \in \mathbb{R}^*$, $N^k(f, y, \cdot) : \mathbb{R} \to L_1(\mathbb{R})$ and $D(f, y, \cdot) : \mathbb{R} \to \mathbb{R}$ given by

$$N^k(f, y, \eta)[x'] \equiv (x')^k \int q(x' \mid x) p(y \mid x, \eta) f(x) dx, \quad D(f, y, \eta) \equiv \int p(y \mid x, \eta) f(x) dx.$$

Note that $D(f, y, \cdot)$ is a real function such that $D(f, y, \cdot) > 0$ on $\mathbb{R}$. The idea now is to recursively apply the product rule for Fréchet derivatives to show that if $N^k(f, y, \cdot) : \mathbb{R} \to L_1(\mathbb{R})$ and $D(f, y, \cdot) : \mathbb{R} \to \mathbb{R}$ are n-times differentiable on $\mathbb{R}$ with nth order Fréchet derivatives $N^k_{(n)}(f, y, \cdot) : \mathbb{R} \to L_1(\mathbb{R})$ and $D_{(n)}(f, y, \cdot) : \mathbb{R} \to \mathbb{R}$, respectively, then so is $\Phi^k(f, y, \cdot)$.

We now show that both $N^k(f, y, \cdot)$ and $D(f, y, \cdot)$ are n-times differentiable on $\mathbb{R}$. We start with $D(f, y, \cdot)$. Since $p(y \mid x, \eta) = \exp(-\eta x) p_\epsilon(\exp(-\eta x) y)$ with $p_\epsilon \in \mathcal{C}^n(\mathbb{R})$, using Leibnitz's formula we obtain

$$\begin{aligned}p_{(n)}(y \mid x, \eta) &\equiv \frac{\partial^n p(y \mid x, \eta)}{\partial \eta^n} \\ &= \sum_{j=0}^{n} \frac{n!}{(n-j)! j!} [p_\epsilon(\exp(-\eta x) y)]_{(j)} (-x)^{n-j} \exp(-\eta x),\end{aligned}$$

where the jth order derivative of the composite function $p_\epsilon(\exp(-\eta x) y)$ can be obtained from Faà di

Bruno's formula,

$$[p_\epsilon(\exp(-\eta x)y)]_{(j)} = \sum \frac{j!}{b_1!\dots b_j!} p_\epsilon^{(r)}(\exp(-\eta x)y) \prod_{i=1}^{j} \left(\frac{y(-x)^i \exp(-\eta x)}{i!} \right)^{b_i},$$

where the sum ranges over all different solutions in nonnegative integers $(b_1, \dots, b_j)$ of $b_1+2b_2+\dots+jb_j = j$ and where r is defined as $r = b_1 + \dots + b_j$. Thus,

$$[p_\epsilon(\exp(-\eta x)y)]_{(j)} = \sum \frac{j!}{b_1!\dots b_j!} p_\epsilon^{(r)}(\exp(-\eta x)y) [y\exp(-\eta x)]^r (-x)^j \frac{1}{1!^{b_1}\dots j!^{b_j}},$$

which combined with the above yields

$$\begin{aligned} & p_{(n)}(y \mid x, \eta) \\ & = (-x)^n \exp(-\eta x) \sum_{j=0}^{n} \frac{n!}{(n-j)!j!} \sum \frac{j!}{b_1!\dots b_j!} p_\epsilon^{(r)}(\exp(-\eta x)y) [y\exp(-\eta x)]^r \frac{1}{1!^{b_1}\dots j!^{b_j}}. \end{aligned} \tag{17}$$

In particular, for $n = 1$ and $n = 2$, the above formula correctly yields

$$p_{(1)}(y \mid x, \eta) = -x\exp(-\eta x) \left[p_\epsilon(\exp(-\eta x)y) + p_\epsilon^{(1)}(\exp(-\eta x)y) y \exp(-\eta x) \right]$$

which was the expression found in Lemma 1, and

$$\begin{aligned} & p_{(2)}(y \mid x, \eta) \\ & = x^2 \exp(-\eta x) \left[p_\epsilon(\exp(-\eta x)y) + 3p_\epsilon^{(1)}(\exp(-\eta x)y) y \exp(-\eta x) + p_\epsilon^{(2)}(\exp(-\eta x)y) [y\exp(-\eta x)]^2 \right]. \end{aligned} \tag{18}$$

It follows from (17), that for every $y \in \mathbb{R}^*$,

$$\left| p_{(n)}(y \mid x, \eta) \right| \leqslant \frac{|x|^n}{|y|} \sum_{j=0}^{n} c_j \sup_{u\in\mathbb{R}} |u^{j+1} p_\epsilon^{(j)}(u)|,$$

where $c_0, \dots, c_n$ are positive finite constants. So if

$$\int_{\mathbb{R}} |x|^n f(x) dx < \infty,$$

Lebesgue dominated convergence theorem implies that $D(f, y, \cdot)$ is n-times differentiable on $\mathbb{R}$ with

$$D_{(n)}(f, y, \eta) = \int p_{(n)}(y \mid x, \eta) f(x) dx. \tag{19}$$

To establish the nth order Fréchet differentiability of $N^k(f, y, \cdot)$, we consider the limit as $h \to 0$ of

$$\begin{aligned}
&\frac{1}{|h|}\bigg\| \int (\cdot)^k q(\cdot \mid x) p_{(n-1)}(y \mid x, \eta + h) f(x) dx - \int (\cdot)^k q(\cdot \mid x) p_{(n-1)}(y \mid x, \eta) f(x) dx \\
&\qquad\qquad\qquad\qquad - h \int (\cdot)^k q(\cdot \mid x) p_{(n)}(y \mid x, \eta) f(x) dx \bigg\|_1 \\
&\leqslant \int\int |x'|^k q(x' \mid x) \frac{\left|p_{(n-1)}(y \mid x, \eta + h) - p_{(n-1)}(y \mid x, \eta) - h p_{(n)}(y \mid x, \eta)\right|}{|h|} f(x) dx dx' \equiv I_k. \qquad (20)
\end{aligned}$$

Consider first the value $k = 0$; then from (20) we have

$$\begin{aligned}
I_0 &= \int \frac{\left|p_{(n-1)}(y \mid x, \eta + h) - p_{(n-1)}(y \mid x, \eta) - h p_{(n)}(y \mid x, \eta)\right|}{|h|} f(x) dx \\
&= \int \left|p_{(n)}(y \mid x, \eta + h^*) - p_{(n)}(y \mid x, \eta)\right| f(x) dx,
\end{aligned}$$

for some $h^* \in (0, h)$, where the second equality comes from the mean value theorem applied to $p_{(n-1)}(y \mid x, \cdot)$, which holds because $p(y \mid x, \cdot) \in \mathcal{C}^n(\mathbb{R})$. Now,

$$\begin{aligned}
\left|p_{(n)}(y \mid x, \eta + h^*) - p_{(n)}(y \mid x, \eta)\right| &\leqslant \left|p_{(n)}(y \mid x, \eta + h^*)\right| + \left|p_{(n)}(y \mid x, \eta)\right| \\
&\leqslant 2\frac{|x|^n}{|y|} \sum_{j=0}^{n} c_j \sup_{u \in \mathbb{R}} |u^{j+1} p_\epsilon^{(j)}(u)|,
\end{aligned}$$

where as before $c_0, \ldots, c_n$ are positive finite constants that come from (17). So if

$$\int_{\mathbb{R}} |x|^n f(x) dx < \infty,$$

we can again use the Lebesgue dominated convergence theorem to pass the limit inside the integral and establish that $N^0(f, y, \cdot)$ is n-times Fréchet differentiable on $\mathbb{R}$ with nth order derivative

$$N^0_{(n)}(f, y, \eta)[x'] = \int q(x' \mid x) p_{(n)}(y \mid x, \eta) f(x) dx. \qquad (21)$$

Now consider the case $k \geqslant 1$ in (12):

$$\begin{aligned}
I_k &\leqslant \int\int \left(|x' - \lambda x| + |\lambda x|\right)^k q(x' \mid x) \frac{\left|p_{(n-1)}(y \mid x, \eta + h) - p_{(n-1)}(y \mid x, \eta) - h p_{(n)}(y \mid x, \eta)\right|}{|h|} f(x) dx dx' \\
&\leqslant C_k \int \left[\frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) + |\lambda x|^k\right] \frac{\left|p_{(n-1)}(y \mid x, \eta + h) - p_{(n-1)}(y \mid x, \eta) - h p_{(n)}(y \mid x, \eta)\right|}{|h|} f(x) dx,
\end{aligned}$$

where as before C_k is a constant ($1 \leqslant C_k < \infty$) such that $(|a| + |b|)^k \leqslant C_k \left(|a|^k + |b|^k\right)$ with $a, b \in \mathbb{R}$. Then, using a reasoning similar to that above, if both $\int |x|^n f(x) dx < \infty$ and $\int |x|^{k+n} f(x) dx < \infty$, it follows that $N^k(f, y, \cdot)$ is n-times Fréchet differentiable on $\mathbb{R}$ with nth order derivative

$$N^k_{(n)}(f, y, \eta)[x'] = (x')^k \int q(x' \mid x) p_{(n)}(y \mid x, \eta) f(x) dx, \quad k \geqslant 1. \tag{22}$$

Combining both results (21) and (22) then gives

$$N^k_{(n)}(f, y, \eta)[x'] = (x')^k \int q(x' \mid x) p_{(n)}(y \mid x, \eta) f(x) dx, \quad k \geqslant 0. \tag{23}$$

Finally, combining (19) and (23) then establishes that for $k \geqslant 0$, $\Phi^k(f, y, \cdot)$ is n-times Fréchet differentiable on $\mathbb{R}$. The nth order Fréchet derivative $\Phi^k_{(n)}(f, y, \cdot)$ of $\Phi^k(f, y, \cdot)$ can be computed recursively as:

$$\Phi^k_{(n)}(f, y, \eta) = \frac{1}{D(f, y, \eta)} \left(N^k_{(n)}(f, y, \eta) - n! \sum_{j=1}^{n} \frac{D_{(n+1-j)}(f, y, \eta)}{(n+1-j)!} \frac{\Phi^k_{(j-1)}(f, y, \eta)}{(j-1)!} \right).$$

Combining the above with (19) and (23) gives:

$$\begin{aligned}\Phi^k_{(n)}(f, y, \eta) = \frac{1}{\int p(y \mid x, \eta) f(x) dx} \Bigg(&(x')^k \int q(x' \mid x) p_{(n)}(y \mid x, \eta) f(x) dx \\ &- n! \sum_{j=1}^{n} \frac{\int p_{(n+1-j)}(y \mid x, \eta) f(x) dx}{(n+1-j)!} \frac{\Phi^k_{(j-1)}(f, y, \eta)}{(j-1)!} \Bigg),\end{aligned}$$

with $p_{(n)}(y \mid x, \eta)$ as given in (17). $\square$

Proof of Theorem 1. The proof is by induction. In what follows, $p_t(\eta) = p_t(\eta, \overline{p})$ and $p_t^{[1]}(\eta) = p_t^{[1]}(\eta, \overline{p})$, i.e. we drop the reference to the prior (the latter being the stationary density $\overline{p}$ of x_t).

Recursion $t = 1$ We have:

$$p_1(\eta) = \phi(\overline{p}, y_1, \eta).$$

Now, consider the following

$$p_1^{[1]}(\eta) = \phi(\overline{p}, y_1, 0) + \phi_\eta(\overline{p}, y_1, 0)\eta.$$

The idea is to show that $p_1^{[1]}(\eta) \in L_1(\mathbb{R})$, and that

$$\|p_1(\eta) - p_1^{[1]}(\eta)\|_1 = o(|\eta|) \quad a.s.$$
$$\int \left| x^k \left(p_1(\eta)[x] - p_1^{[1]}(\eta)[x] \right) \right| dx = o(|\eta|) \quad a.s. \quad \text{for every } k \geqslant 1,$$

so that we can call $p_1^{[1]}(\eta)$ a first order approximation to $p_1(\eta)$.

Note that for any prior density p_0

$$\phi(p_0, y_1, 0)[x] = \frac{\int q(x \mid x_0) p_\epsilon(y_1) p_0(x_0) dx_0}{\int p_\epsilon(y_1) p_0(x_0) dx_0} = \int q(x \mid x_0) p_0(x_0) dx_0,$$

so $\phi(\cdot, y_1, 0)$ is a linear operator that is independent of y_1. We can write

$$\phi(p_0, y_1, 0) = Lp_0 \quad \text{with} \quad Lp_0(x) \equiv \int q(x \mid x_0) p_0(x_0) dx_0.$$

Notice that if $p_0 = \overline{p}$, i.e. if we start with the unconditional distribution of x_t, then

$$L\overline{p} = \overline{p}.$$

For the second term, note that from Lemma 1 we have:

$$\phi_\eta(p_0, y_1, \eta)[x] = \frac{\int q(x \mid x_0) p_\eta(y_1 \mid x_0, \eta) p_0(x_0) dx_0}{\int p(y_1 \mid x_0, \eta) p_0(x_0) dx_0} - \frac{\left[\int q(x \mid x_0) p(y_1 \mid x_0, \eta) p_0(x_0) dx_0\right]\left[\int p_\eta(y_1 \mid x_0, \eta) p_0(x_0) dx_0\right]}{\left[\int p(y_1 \mid x_0, \eta) p_0(x_0) dx_0\right]^2}$$

so that

$$\begin{aligned} \phi_\eta(p_0, y_1, 0)[x] &= \left[-\int q(x \mid x_0) x_0 p_0(x_0) dx_0\right] \psi_1(y_1) + \left[\int q(x \mid x_0) p_0(x_0) dx_0\right]\left[\int x_0 p_0(x_0) dx_0\right] \psi_1(y_1) \\ &= -\psi_1(y_1)\left\{\int q(x \mid x_0) x_0 p_0(x_0) dx_0 - \left[\int q(x \mid x_0) p_0(x_0) dx_0\right]\left[\int x_0 p_0(x_0) dx_0\right]\right\}. \end{aligned}$$

When the prior is chosen as $p_0 = \overline{p}$, then

$$\begin{gathered} \int q(x \mid x_0) \overline{p}(x_0) dx_0 = \overline{p}(x) \\ \int x_0 \overline{p}(x_0) dx_0 = 0 \\ \int q(x \mid x_0) x_0 \overline{p}(x_0) dx_0 = \lambda x \overline{p}(x) \end{gathered}$$

This is because

$$\begin{aligned} \int q(x \mid x_0) x_0 \overline{p}(x_0) dx_0 &= \int x_0 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \lambda x_0)^2}{2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_0^2}{2\sigma^2}\right) dx_0 \\ &= \int \frac{1}{\sqrt{2\pi(1+\lambda^2\sigma^2)}} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1+\lambda^2\sigma^2}}} x_0 \exp\left[\frac{\left(x_0 - \lambda\frac{\sigma^2}{1+\lambda^2\sigma^2}x\right)^2}{2\frac{\sigma^2}{1+\lambda^2\sigma^2}}\right] \exp\left[-\frac{x^2}{2(1+\lambda^2\sigma^2)}\right] dx_0 \\ &= \left[\lambda\frac{\sigma^2}{1+\lambda^2\sigma^2}x\right] \frac{1}{\sqrt{2\pi(1+\lambda^2\sigma^2)}} \exp\left[-\frac{x^2}{2(1+\lambda^2\sigma^2)}\right] \\ &= \lambda x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \end{aligned}$$

where we have used the fact that $1 + \lambda^2\sigma^2 = \sigma^2$.

Combining all of the above, we get:

$$p_1^{[1]}(\eta)[x] = \overline{p}(x) - \eta\psi_1(y_1)\lambda x \overline{p}(x) = \overline{p}(x)\left[1 - \eta\psi_1(y_1)\lambda x\right].$$

Notice that for every $y_1 \in \mathbb{R}$,

$$\int |p_1^{[1]}(\eta)[x]|dx = \int |1 - \eta\psi_1(y_1)\lambda x|\overline{p}(x)dx < \infty, \quad \text{so} \quad p_1^{[1]}(\eta) \in L_1(\mathbb{R}).$$

Moreover, $\int p_1^{[1]}(\eta)[x]dx = 1$. In addition, since $\int |x|^k\overline{p}(x)dx < \infty$ for any $k \geqslant 1$, for any $y_1 \in \mathbb{R}^*$, by the Fréchet differentiability of $\phi(\overline{p}, y_1, \cdot) : \mathbb{R} \to L_1(\mathbb{R})$ established in Lemma 1 (case $k = 0$) we have:

$$\|\phi(\overline{p}, y_1, \eta) - \phi(\overline{p}, y_1, 0) - \phi_\eta(\overline{p}, y_1, 0)\eta\|_1 = o(|\eta|) \tag{24}$$

So letting

$$A_{1,0} = 0, \quad A_{1,1} = \lambda\left[A_{1,0} - \psi_1(y_1)\right]$$

we have:

$$\|p_1(\eta) - p_1^{[1]}(\eta)\|_1 = o(|\eta|) \quad a.s. \quad \text{where} \quad p_1^{[1]}(\eta)[x] = \overline{p}(x)\left[1 + A_{1,1}\eta x\right].$$

It is clear that $\int |x^k p_1^{[1]}(\eta)[x]|dx < \infty$ for every $k \geqslant 1$, and

$$\begin{aligned}\int \left|x^k\left(p_1(\eta)[x] - p_1^{[1]}(\eta)[x]\right)\right| dx &= \int \left|x^k\phi(\overline{p}, y_1, \eta)[x] - x^k\phi(\overline{p}, y_1, 0)[x] - x^k\phi_\eta(\overline{p}, y_1, 0)[x]\eta\right| dx \\ &= \left\|\Phi^k(\overline{p}, y_1, \eta) - \Phi^k(\overline{p}, y_1, 0) - \Phi^k_\eta(\overline{p}, y_1, 0)\eta\right\|_1 \\ &= o(|\eta|)\end{aligned}$$

where the last equality uses the differentiability of $\Phi^k(\overline{p}, y_1, \eta)$ established in Lemma 1. This completes the proof of the result for $t = 1$.

Recursion any t. Start with $p_t^{[1]}(\eta)$ with $p_t^{[1]}(\eta)[x] \equiv \overline{p}(x)\left[1 + A_{1,t}\eta x\right]$, and

$$A_{1,t} = \lambda\left[A_{1,t-1} - \psi_1(y_t)\right],$$

where we know that

$$\begin{aligned}\|p_t(\eta) - p_t^{[1]}(\eta)\|_1 = o(|\eta|) \quad a.s. \\ \int \left|x^k\left(p_t(\eta)[x] - p_t^{[1]}(\eta)[x]\right)\right| dx = o(|\eta|) \quad a.s. \quad \text{for every } k \geqslant 1.\end{aligned} \tag{25}$$

Then let $p_{t+1}^{[1]}(\eta)[x] = \overline{p}(x)\left[1 + A_{1,t+1}\eta x\right]$ with

$$A_{1,t+1} = \lambda\left[A_{1,t} - \psi_1(y_{t+1})\right],$$

and show that the same properties as in (25) hold at $t + 1$.

Write $p_t(\eta) = p_t^{[1]}(\eta) + \rho_t$ where $\|\rho_t\|_1 = o(|\eta|)$, and let

$$\begin{aligned}\widetilde{p}_{t+1}(\eta) &\equiv \phi(p_t(\eta), y_{t+1}, 0) + \phi_\eta(p_t(\eta), y_{t+1}, 0)\eta \\ &= \phi(p_t^{[1]}(\eta) + \rho_t, y_{t+1}, 0) + \phi_\eta(p_t^{[1]}(\eta) + \rho_t, y_{t+1}, 0)\eta.\end{aligned}$$

As before:

$$\begin{aligned}\phi(p_t(\eta), y_{t+1}, 0)[x] &= \int q(x \mid x_t)p_t(\eta)[x_t]dx_t, \\ \phi_\eta(p_t(\eta), y_{t+1}, 0)[x] &= -\psi_1(y_{t+1})\left\{\int q(x \mid x_t)x_t p_t(\eta)[x_t]dx_t - \left[\int q(x \mid x_t)p_t(\eta)[x_t]dx_t\right]\left[\int x_t p_t(\eta)[x_t]dx_t\right]\right\}\end{aligned}$$

so

$$\begin{aligned}\phi(p_t^{[1]}(\eta) + \rho_t, y_{t+1}, 0)[x] &= \int q(x \mid x_t)\overline{p}(x_t)\left[1 + A_{1,t}\eta x_t\right]dx_t + \int q(x \mid x_t)\rho_t(x_t)dx_t \\ &= \int q(x \mid x_t)\overline{p}(x_t)dx_t + \int q(x \mid x_t)\overline{p}(x_t)A_{1,t}\eta x_t dx_t + \int q(x \mid x_t)\rho_t(x_t)dx_t \\ &= \overline{p}(x) + A_{1,t}\eta\int q(x \mid x_t)\overline{p}(x_t)x_t dx_t + \int q(x \mid x_t)\rho_t(x_t)dx_t \\ &= \overline{p}(x)\left[1 + \lambda A_{1,t}\eta x\right] + \int q(x \mid x_t)\rho_t(x_t)dx_t\end{aligned}$$

and

$$\begin{aligned}\phi_\eta(p_t^{[1]}(\eta) + \rho_t, y_{t+1}, 0)[x] = -\psi_1(y_{t+1})\Bigg\{&\int q(x \mid x_t)x_t\{p_t^{[1]}(\eta)[x_t] + \rho_t(x_t)\}dx_t \\ &- \left[\int q(x \mid x_t)\{p_t^{[1]}(\eta)[x_t] + \rho_t(x_t)\}dx_t\right]\left[\int x_t\{p_t^{[1]}(\eta)[x_t] + \rho_t(x_t)\}dx_t\right]\Bigg\}.\end{aligned}$$

Now we separately consider the three terms above. First

$$\begin{aligned}&\int q(x \mid x_t)x_t\{p_t^{[1]}(\eta)[x_t] + \rho_t(x_t)\}dx_t \\ &\qquad = \int q(x \mid x_t)x_t p_t^{[1]}(\eta)[x_t]dx_t + \int q(x \mid x_t)x_t\rho_t(x_t)dx_t \\ &\qquad = \int q(x \mid x_t)\overline{p}(x_t)x_t dx_t + A_{1,t}\eta\int q(x \mid x_t)\overline{p}(x_t)x_t^2 dx_t + \int q(x \mid x_t)x_t\rho_t(x_t)dx_t \\ &\qquad = \lambda x\overline{p}(x) + A_{1,t}\eta[1 + \lambda^2 x^2]\overline{p}(x) + \int q(x \mid x_t)x_t\rho_t(x_t)dx_t.\end{aligned}$$

Next, as from our calculations above

$$\int q(x \mid x_t)\{p_t^{[1]}(\eta)[x_t] + \rho_t(x_t)\}dx_t = \overline{p}(x) + \lambda A_{1,t}\eta x\overline{p}(x) + \int q(x \mid x_t)\rho_t(x_t)dx_t,$$

and finally

$$\int x_t\{p_t^{[1]}(\eta)[x_t] + \rho_t(x_t)\}dx_t = A_{1,t}\eta \int \overline{p}(x_t)x_t^2 dx_t + \int x_t \rho_t(x_t)dx_t$$
$$= A_{1,t}\eta\sigma^2 + \int x_t\rho_t(x_t)dx_t.$$

In the above, we have used the facts that:

$$\int q(x \mid x_t)\overline{p}(x_t)x_t dx_t = \lambda x \overline{p}(x)$$
$$\int \overline{p}(x_t)x_t^2 dx_t = \sigma^2$$
$$\int q(x \mid x_t)\overline{p}(x_t)x_t^2 dx_t = \left[1 + \lambda^2 x^2\right]\overline{p}(x).$$

Now, combine all of the above to get

$$\phi_\eta(p_t^{[1]}(\eta) + \rho_t, y_{t+1}, 0)[x] = -\psi_1(y_{t+1})\lambda x\overline{p}(x) + R_{t+1}(x)$$

where

$$\begin{aligned} R_{t+1}(x) \equiv -\psi_1(y_{t+1})\Bigg\{ & A_{1,t}\eta[1+\lambda^2x^2]\overline{p}(x) + \int q(x \mid x_t)x_t\rho_t(x_t)dx_t \\ & - \left[\overline{p}(x) + \lambda A_{1,t}\eta x\overline{p}(x) + \int q(x \mid x_t)\rho_t(x_t)dx_t\right]\left[A_{1,t}\eta\sigma^2 + \int x_t\rho_t(x_t)dx_t\right]\Bigg\} \\ = -\psi_1(y_{t+1})\Bigg\{ & A_{1,t}\eta[1-\sigma^2+\lambda^2x^2]\overline{p}(x) + \int q(x \mid x_t)x_t\rho_t(x_t)dx_t - \overline{p}(x)\int x_t\rho_t(x_t)dx_t \\ & - \left[\lambda A_{1,t}\eta x\overline{p}(x) + \int q(x \mid x_t)\rho_t(x_t)dx_t\right]\left[A_{1,t}\eta\sigma^2 + \int x_t\rho_t(x_t)dx_t\right]\Bigg\} \end{aligned} \tag{26}$$

Note that

$$\begin{aligned} \|R_{t+1}\|_1 \leqslant |\psi_1(y_{t+1})|\Bigg\{ & |A_{1,t}\eta|(1+\lambda^2\sigma^2) + \int \left|\int q(x \mid x_t)x_t\rho_t(x_t)dx_t\right| dx + \left|\int x_t\rho_t(x_t)dx_t\right| \\ & + \left[|A_{1,t}\eta\sigma^2| + \left|\int x_t\rho_t(x_t)dx_t\right|\right]\left[|\lambda A_{1,t}\eta|\int |x|\overline{p}(x)dx + \int\left|\int q(x \mid x_t)\rho_t(x_t)dx_t\right|dx\right]\Bigg\}, \end{aligned}$$

and

$$\int \left|\int q(x \mid x_t)x_t\rho_t(x_t)dx_t\right| dx \leqslant \int\int q(x \mid x_t)|x_t\rho_t(x_t)|dx_t dx = \int |x_t\rho_t(x_t)|dx_t$$
$$\left|\int x_t\rho_t(x_t)dx_t\right| \leqslant \int |x_t\rho_t(x_t)|dx_t$$
$$\int \left|\int q(x \mid x_t)\rho_t(x_t)dx_t\right| dx \leqslant \int\int q(x \mid x_t)|\rho_t(x_t)|dx_t dx = \int |\rho_t(x_t)|dx_t$$

so using (25) it follows that $\|R_{t+1}\|_1 = O(|\eta|)$. Putting everything together

$$\begin{aligned}\widetilde{p}_{t+1}(\eta)[x] &= \overline{p}(x)\left[1+\lambda\left(A_{1,t}-\psi_1(y_{t+1})\right)\eta x\right] + \int q(x \mid x_t)\rho_t(x_t)dx_t + \eta R_{t+1}(x) \\ &= p^{[1]}_{t+1}(\eta)[x] + \int q(x \mid x_t)\rho_t(x_t)dx_t + \eta R_{t+1}(x).\end{aligned}$$

Now let

$$\widetilde{R}_{t+1}(x) \equiv \int q(x \mid x_t)\rho_t(x_t)dx_t + \eta R_{t+1}(x). \tag{27}$$

Then,

$$\begin{aligned}\|p_{t+1}(\eta) - p^{[1]}_{t+1}(\eta)\|_1 &= \|p_{t+1}(\eta) - \widetilde{p}_{t+1}(\eta) - \widetilde{R}_{t+1}\|_1 \\ &\leqslant \|p_{t+1}(\eta) - \widetilde{p}_{t+1}(\eta)\|_1 + \|\widetilde{R}_{t+1}\|_1 \\ &= \|\phi(p_t(\eta), y_{t+1}, \eta) - \phi(p_t(\eta), y_{t+1}, 0) - \phi_\eta(p_t(\eta), y_{t+1}, 0)\eta\|_1 + \|\widetilde{R}_{t+1}\|_1 \\ &= o(|\eta|) + \|\widetilde{R}_{t+1}\|_1\end{aligned}$$

where we have used the Fréchet differentiability of $\phi(p_t(\eta), y_{t+1}, \cdot)$ at $\eta = 0$. For the second term, notice that

$$\begin{aligned}\|\widetilde{R}_{t+1}\|_1 &\leqslant \int \left|\int q(x \mid x_t)\rho_t(x_t)dx_t\right| dx + |\eta|\|R_{t+1}\|_1 \\ &\leqslant \|\rho_t\|_1 + |\eta|\|R_{t+1}\|_1 \\ &= o(|\eta|),\end{aligned}$$

where we have used the fact that $\|R_{t+1}\|_1 = O(|\eta|)$. Thus

$$\|p_{t+1}(\eta) - p^{[1]}_{t+1}(\eta)\|_1 = o(|\eta|) \quad a.s. \tag{28}$$

It is clear that for every $k \geqslant 1$, $\int |x^k p^{[1]}_{t+1}(\eta)[x]|dx < \infty$, so it remains to show that

$$\int \left|x^k \left(p_{t+1}(\eta)[x] - p^{[1]}_{t+1}(\eta)[x]\right)\right| dx = o(|\eta|) \quad a.s. \quad \text{for every } k \geqslant 1 \tag{29}$$

Following the same reasoning as above, the property (29) follows from the Fréchet differentiability of $\Phi^k(p_t(\eta), y_{t+1}, \cdot)$ established in Lemma 1, provided we can show that

$$\int \left|x^k \widetilde{R}_{t+1}(x)\right| dx = o(|\eta|) \quad a.s. \quad \text{for every } k \geqslant 1.$$

From (27), we have

$$\begin{aligned}\|(\cdot)^k \widetilde{R}_{t+1}(\cdot)\|_1 &\leqslant \int |x|^k \left| \int q(x \mid x_t)\rho_t(x_t)dx_t \right| dx + |\eta| \|(\cdot)^k R_{t+1}(\cdot)\|_1 \\ &\leqslant \|(\cdot)^k \rho_t(\cdot)\|_1 + |\eta| \|(\cdot)^k R_{t+1}(\cdot)\|_1,\end{aligned}$$

where the first term on the right-hand side of the last inequality is $o(|\eta|)$ by (16). We now establish that moreover $\|(\cdot)^k \widetilde{R}_{t+1}(\cdot)\|_1 = O(|\eta|)$. From (26), we have:

$$\begin{aligned}&\int |x^k R_{t+1}(x)|dx \leqslant \\ &|\psi_1(y_{t+1})| \Bigg\{ |A_{1,t}\eta| \int |x^k [1-\sigma^2+\lambda^2 x^2]| \overline{p}(x)dx + \int \left| x^k \int q(x \mid x_t) x_t \rho_t(x_t)dx_t \right| dx + \left| \int x_t \rho_t(x_t)dx_t \right| \\ &\quad + \left[|A_{1,t}\eta\sigma^2| + \left| \int x_t \rho_t(x_t)dx_t \right| \right] \left[|\lambda A_{1,t}\eta| \int |x|^{k+1} \overline{p}(x)dx + \left| x^k \int q(x \mid x_t)\rho_t(x_t)dx_t \right| dx \right] \Bigg\}.\end{aligned}$$

Now, using the fact that for every $k \geqslant 1$, there exists a $1 \leqslant C_k < \infty$, such that $(|a|+|b|)^k \leqslant C_k \left(|a|^k + |b|^k \right)$ $(a, b \in \mathbb{R})$, we obtain

$$\begin{aligned}\int \left| x^k \int q(x \mid x_t) x_t \rho_t(x_t)dx_t \right| dx &\leqslant \int\int |x^k| q(x \mid x_t) |x_t \rho_t(x_t)| dx_t dx \\ &\leqslant \int\int (|x-\lambda x_t| + |\lambda x_t|)^k q(x \mid x_t) |x_t \rho_t(x_t)| dx_t dx \\ &\leqslant C_k \int\int \left(|x-\lambda x_t|^k + |\lambda x_t|^k \right) q(x \mid x_t) |x_t \rho_t(x_t)| dx_t dx \\ &\leqslant C_k \left\{ \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) \int |x_t \rho_t(x_t)| dx_t + |\lambda|^k \int |x_t^{k+1} \rho_t(x_t)| dx_t \right\} \\ &= o(|\eta|)\end{aligned}$$

where the last equality uses (16). The latter also implies that $\left| \int x_t \rho_t(x_t)dx_t \right| \leqslant \int |x_t \rho_t(x_t)| dx_t = o(|\eta|)$, and that

$$\begin{aligned}\int \left| x^k \int q(x \mid x_t)\rho_t(x_t)dx_t \right| dx &\leqslant \int\int |x^k| q(x \mid x_t) |\rho_t(x_t)| dx_t dx \\ &\leqslant \int\int (|x-\lambda x_t| + |\lambda x_t|)^k q(x \mid x_t) |\rho_t(x_t)| dx_t dx \\ &\leqslant C_k \int\int \left(|x-\lambda x_t|^k + |\lambda x_t|^k \right) q(x \mid x_t) |\rho_t(x_t)| dx_t dx \\ &\leqslant C_k \left\{ \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) \int |\rho_t(x_t)| dx_t + |\lambda|^k \int |x_t^k \rho_t(x_t)| dx_t \right\} \\ &= o(|\eta|).\end{aligned}$$

Combining all of the above yields $\|(\cdot)^k \widetilde{R}_{t+1}(\cdot)\|_1 = O(|\eta|)$. This establishes the result at $t+1$. $\square$

Proof of Corollary 1. As before, we drop any reference to the prior $\overline{p}$. By triangle inequality,

$$\|p_t(\eta) - p_t^{[1]}(\varsigma)\|_1 \leqslant \|p_t(\eta) - p_t^{[1]}(\eta)\|_1 + \|p_t^{[1]}(\varsigma) - p_t^{[1]}(\eta)\|_1,$$

where $p_t^{[1]}(\eta)$ is the first-order perturbation approximation from Theorem 1. As is shown in Theorem 1, the first term $\|p_t(\eta) - p_t^{[1]}(\eta)\|_1$ is $o(|\eta|)$ on $\mathbb{R}$. For the second term, note that

$$\begin{aligned} p_t^{[1]}(\varsigma)[x] - p_t^{[1]}(\eta)[x] &= \overline{p}(x)xA_{1,t}\left[\frac{S(\eta)}{S'(0)} - \eta\right] \\ &= \overline{p}(x)xA_{1,t}\left[\frac{1}{2}\frac{S''(\xi)}{S'(0)}\eta^2\right], \end{aligned}$$

where the second equality follows by the mean-value theorem, $S(\eta) = S(0) + S'(0)\eta + 1/2S''(\xi)\eta^2 = S'(0)\eta + 1/2S'(\xi)\eta^2$, and ξ belongs to the interval with endpoints 0 and η. Therefore,

$$\|p_t^{[1]}(\varsigma) - p_t^{[1]}(\eta)\|_1 = |A_{1,t}|\frac{1}{2}\frac{|S''(\xi)|}{|S'(0)|}\sqrt{\frac{2}{\pi}}\eta^2 = o(|\eta|) \quad a.s.$$

since as η goes to zero, $|S''(\xi)|$ converges to $|S''(0)| < \infty$. Finally, note that any $o(|\eta|)$ sequence is also $o(|\varsigma|)$, because $\lim_{\eta\to 0} S(\eta)/\eta = S'(0)$, which is finite. □

Proof of Lemma 4. Under Assumption 3, we have for any $y \in \mathbb{R}^* \equiv \mathbb{R} \setminus \{0\}$,

$$p(y \mid x, \eta) = \frac{1}{|y|}\left|\exp(-\eta x)yp_\epsilon(\exp(-\eta x)y)\right|,$$

so for any $y \in \mathbb{R}^*$,

$$\sup_{x\in\mathbb{R}} p(y \mid x, \eta) \leqslant \frac{1}{|y|}\sup_{u\in\mathbb{R}}|up_\epsilon(u)| < \infty.$$

Under Assumption 2, there exists a function $V : \mathbb{R} \to [1, +\infty)$ such that

$$S_V \equiv \sup_{x\in\mathbb{R}} \frac{\int_{\mathbb{R}} q(x' \mid x)V(x')dx'}{V(x)} < \infty \quad \text{and} \quad \lim_{|x|\to\infty} \frac{\int_{\mathbb{R}} q(x' \mid x)V(x')dx'}{V(x)} = 0. \tag{30}$$

Since $q(x' \mid x) = \frac{1}{\sqrt{2\pi}}\exp(-(x' - \lambda x)^2/2)$, it suffices to take $V(x) = \exp(\gamma|x|)$ with $\gamma > 0$, since then

$$\begin{aligned} \int_{\mathbb{R}} q(x' \mid x)V(x')dx' &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left[-\frac{(x' - \lambda x)^2}{2}\right]\exp\left[\gamma|x'|\right]dx' \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left[-\frac{u^2}{2}\right]\exp\left[\gamma|u + \lambda x|\right]du \\ &= \exp\left[-\gamma\lambda x\right]\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\lambda x} \exp\left[-\frac{u^2 + 2\gamma u}{2}\right]du + \exp\left[\gamma\lambda x\right]\frac{1}{\sqrt{2\pi}}\int_{-\lambda x}^{+\infty} \exp\left[-\frac{u^2 - 2\gamma u}{2}\right]du \\ &= \exp[\gamma^2/2]\left\{\exp\left[-\gamma\lambda x\right]\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\lambda x+\gamma} \exp\left[-\frac{v^2}{2}\right]dv + \exp\left[\gamma\lambda x\right]\frac{1}{\sqrt{2\pi}}\int_{-\lambda x-\gamma}^{+\infty} \exp\left[-\frac{v^2}{2}\right]dv\right\} \\ &= \exp[\gamma^2/2]\left\{\exp\left[-\gamma\lambda x\right]\Phi(-\lambda x + \gamma) + \exp\left[\gamma\lambda x\right](1 - \Phi(-\lambda x - \gamma))\right\} \end{aligned}$$

where $\Phi(\cdot)$ denotes the standard normal cdf. Thus

$$\frac{\int_{\mathbb{R}} q(x' \mid x) V(x') dx'}{V(x)} = \exp[\gamma^2/2] \left\{ \frac{\exp(-\gamma\lambda x)}{\exp(\gamma|x|)} \Phi(-\lambda x + \gamma) + \frac{\exp(\gamma\lambda x)}{\exp(\gamma|x|)} \left(1 - \Phi(-\lambda x - \gamma)\right) \right\}. \tag{31}$$

Since $|\lambda| < 1$,

$$\sup_{x\in\mathbb{R}} \frac{\exp(-\gamma\lambda x)}{\exp(\gamma|x|)} < \infty \quad \text{and} \quad \sup_{x\in\mathbb{R}} \frac{\exp(\gamma\lambda x)}{\exp(\gamma|x|)} < \infty$$
$$\lim_{|x|\to\infty} \frac{\exp(-\gamma\lambda x)}{\exp(\gamma|x|)} = 0 \quad \text{and} \quad \lim_{|x|\to\infty} \frac{\exp(\gamma\lambda x)}{\exp(\gamma|x|)} = 0$$

so combining the above with (31) gives (30).

Therefore, since for any $y \in \mathbb{R}^*$, $\sup_{x\in\mathbb{R}} p(y \mid x, \eta) < \infty$, for any $\nu > 0$ one may choose a constant $c > 0$ large enough so that $\Gamma_C(y) \leqslant \nu\Gamma_{\mathbb{R}}(y)$, where for any set $A \subset \mathbb{R}$

$$\Gamma_A(y) \equiv \sup_{x\in A} p(y \mid x, \eta) \frac{\int_{\mathbb{R}} q(x' \mid x) V(x') dx'}{V(x)}, \tag{32}$$

and the set $C \equiv \{x : |x| > c\}$ is a complement of a bounded subset of $\mathbb{R}$, the latter satisfying the local Doeblin property (for state space models with linear Gaussian transition equation, every bounded set is locally Doeblin; see p. 1238 in Douc, Fort, Moulines, and Priouret (2009) for details).

Under Assumptions 3 and 2, assumptions (H1) and (H2) inDouc, Fort, Moulines, and Priouret (2009) are satisfied, with $K = \mathbb{R}^*$. We now check the remaining assumptions in their Theorem 1. We start with their condition (12): for some constant $\delta \in (0, 1)$

$$\liminf_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbb{1}_{\mathbb{R}^*}(y_t) \geqslant (1 + \delta)/2 \quad a.s. \tag{33}$$

Recall that $\{(x_t, y_{t+1})'\}_{t\geqslant 0}$ is a positive recurrent Markov chain with stationary distribution that has a density $\overline{p}(x)p(y \mid x, \eta)$. Moreover, $\{(x_t, y_{t+1})'\}_{t\geqslant 0}$ satisfies the law of large numbers, i.e. for any prior density p_0 and any measurable function $f : \mathbb{R}^2 \to [0, +\infty)$ satisfying $\int_{\mathbb{R}^2} f(x, y)\overline{p}(x)p(y \mid x, \eta)dxdy < \infty$, we have

$$T^{-1} \sum_{t=0}^{T} f(x_t, y_{t+1}) \to \int_{\mathbb{R}^2} f(x, y)\overline{p}(x)p(y \mid x, \eta)dxdy \quad a.s.$$

Letting $f(x_t, y_{t+1}) = \mathbb{1}_{\mathbb{R}^*}(y_{t+1})$, then gives $\liminf_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbb{1}_{\mathbb{R}^*}(y_t) = 1$ a.s. so condition (33) is satisfied.

To check conditions (13) and (14) we use the results of Corollary 5 in Douc, Fort, Moulines, and Priouret (2009). We need

$$\int_{\mathbb{R}^2} [\ln \Gamma_{\mathbb{R}}(y)]_+ \overline{p}(x)p(y \mid x, \eta)dxdy < \infty. \tag{34}$$

For this, combining (32), (30) and the expression for the observation density $p(y \mid x, \eta)$ we have, for any

$y \in \mathbb{R}^*$,

$$\begin{aligned}\Gamma_{\mathbb{R}}(y) &\leqslant \left\{\sup_{x\in\mathbb{R}} [\exp(-\eta x) p_\epsilon(\exp(-\eta x) y)]\right\} S_V \\ &\leqslant S_V \frac{1}{|y|} \sup_{u\in\mathbb{R}} |u p_\epsilon(u)|\end{aligned}$$

so by letting $K \equiv |\ln(S_V \sup_{u\in\mathbb{R}} |u p_\epsilon(u)|)|$, we have $0 < K < \infty$ and

$$[\ln \Gamma_{\mathbb{R}}(y)]_+ \leqslant K + |\ln|y||$$

Thus

$$\begin{aligned}\int_{\mathbb{R}} [\ln \Gamma_{\mathbb{R}}(y)]_+ p(y \mid x, \eta) dy &\leqslant K + \int_{\mathbb{R}} |\ln|y|| \, p(y \mid x, \eta) dy \\ &\leqslant K + |\eta x| + \int_{\mathbb{R}} |\ln|z|| p_\epsilon(z) dz,\end{aligned}$$

provided $\int_{\mathbb{R}} |\ln|z|| p_\epsilon(z) dz < \infty$. To ensure the latter, we impose Assumption 4. Since $\int_{\mathbb{R}} |\eta x| \overline{p}(x) dx < \infty$, condition (34) holds.

To make sure condition (14) in Douc, Fort, Moulines, and Priouret (2009) holds, set $D \equiv [-d, d]$, $0 < d < \infty$; then the set D has a local Doeblin property. Let

$$\Psi_D(y) \equiv \frac{1}{2d} \int_{-d}^{d} p(y \mid x, \eta) dx.$$

We need

$$\int_{\mathbb{R}^2} [\ln \Psi_D(y)]_- \overline{p}(x) p(y \mid x, \eta) dx dy < \infty. \tag{35}$$

By the Jensen inequality, for any $y \in \mathbb{R}^*$,

$$\ln \Psi_D(y) \geqslant \frac{1}{2d} \int_{-d}^{d} \ln p(y \mid x, \eta) dx,$$

so letting $C \equiv \left| \ln \sup_{u\in\mathbb{R}} |u p_\epsilon(u)| \right| < \infty$ we have

$$\begin{aligned}[\ln \Psi_D(y)]_- &\leqslant \frac{1}{2d} \int_{-d}^{d} |\ln p(y \mid x, \eta)| dx \\ &\leqslant \frac{1}{2d} \int_{-d}^{d} \left(|\ln|y|| + \left| \ln \sup_{u\in\mathbb{R}} |u p_\epsilon(u)| \right| \right) dx \\ &= |\ln|y|| + C.\end{aligned}$$

The property in (35) then follows by the same reasoning used to establish the property in (34).

Theorem 1 in Douc, Fort, Moulines, and Priouret (2009) then applies and shows that for any initial distribution p_0 such that for some $\gamma > 0$,

$$\int_{\mathbb{R}} \exp(\gamma|x|) p_0(x) dx < \infty, \tag{36}$$

there exists a positive constant $c > 0$ such that we have

$$\limsup_{t \to \infty} t^{-1} \ln \|p_t(\eta, p_0) - p_t(\eta, \overline{p})\|_1 < -c \quad a.s. \tag{37}$$

$\square$

B Higher order approximations

In this Appendix, we detail the computation of second and third order approximations to the nonlinear filter.

B.1 Second order approximation

Recall from Lemma 3 that the first and second order derivatives of $p(y \mid x, \cdot)$, which we denote by $p_{(1)}(y \mid x, \cdot)$ and $p_{(2)}(y \mid x, \cdot)$, respectively, are given by:

$$p_{(1)}(y \mid x, \eta) = -x \exp(-\eta x) \left[p_\epsilon(\exp(-\eta x) y) + p_\epsilon^{(1)}(\exp(-\eta x) y) y \exp(-\eta x) \right]$$
$$p_{(2)}(y \mid x, \eta) = x^2 \exp(-\eta x) \left[p_\epsilon(\exp(-\eta x) y) + 3 p_\epsilon^{(1)}(\exp(-\eta x) y) y \exp(-\eta x) + p_\epsilon^{(2)}(\exp(-\eta x) y) \left[y \exp(-\eta x) \right]^2 \right].$$

Letting

$$\begin{aligned} \psi_1(y) &\equiv 1 + y \frac{p_\epsilon'(y)}{p_\epsilon(y)} \\ \widetilde{\psi}_2(y) &\equiv 1 + 3y \frac{p_\epsilon'(y)}{p_\epsilon(y)} + y^2 \frac{p_\epsilon''(y)}{p_\epsilon(y)} \end{aligned}$$

we can then write

$$\begin{aligned} p_{(1)}(y \mid x, 0) &= -x p_\epsilon(y) \left[1 + y \frac{p_\epsilon'(y)}{p_\epsilon(y)} \right] = -x p_\epsilon(y) \psi_1(y) \\ p_{(2)}(y \mid x, 0) &= x^2 p_\epsilon(y) \left[1 + 3y \frac{p_\epsilon'(y)}{p_\epsilon(y)} + y^2 \frac{p_\epsilon''(y)}{p_\epsilon(y)} \right] = x^2 p_\epsilon(y) \widetilde{\psi}_2(y). \end{aligned}$$

The construction is by induction. In what follows, we drop the reference to the prior unconditional density $\overline{p}$ and write $p_t(\eta) = p_t(\eta, \overline{p})$ and $p_t^{[2]}(\eta) = p_t^{[2]}(\eta, \overline{p})$.

Recursion $t = 1$ We have:

$$p_1(\eta) = \phi(\overline{p}, y_1, \eta).$$

Now, consider the following

$$p_1^{[2]}(\eta) = \phi(\overline{p}, y_1, 0) + \phi_\eta(\overline{p}, y_1, 0)\eta + \frac{1}{2}\phi_{\eta\eta}(\overline{p}, y_1, 0)\eta^2.$$

As before (see the proof of Theorem 1),

$$\begin{aligned} \phi(\overline{p}, y_1, 0)[x] &= \overline{p}(x) \\ \phi_\eta(\overline{p}, y_1, 0)[x] &= -\phi_1(y_1) \lambda x \overline{p}(x). \end{aligned}$$

Next, for the second-order term:

$$
\begin{aligned}
\phi_{\eta\eta}(\overline{p}, y_1, \eta)[x] = & \frac{\int q(x \mid x_0) p_{\eta\eta}(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0}{\int p(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0} \\
& - 2\frac{\left[\int q(x \mid x_0) p_{\eta}(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]\left[\int p_{\eta}(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]}{\left[\int p(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]^2} \\
& - \frac{\left[\int q(x \mid x_0) p(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]\left[\int p_{\eta\eta}(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]}{\left[\int p(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]^2} \\
& + 2\frac{\left[\int q(x \mid x_0) p(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]\left[\int p_{\eta}(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]^2}{\left[\int p(y_1 \mid x_0, \eta)\overline{p}(x_0) dx_0\right]^3}
\end{aligned}
$$

so that

$$
\begin{aligned}
\phi_{\eta\eta}(\overline{p}, y_1, 0)[x] = & \left[\int q(x \mid x_0) x_0^2 \overline{p}(x_0) dx_0\right] \widetilde{\psi}_2(y_1) - 2\left[\int q(x \mid x_0) x_0 \overline{p}(x_0) dx_0\right]\left[\int x_0 \overline{p}(x_0) dx_0\right] \psi_1^2(y_1) \\
& - \left[\int q(x \mid x_0) \overline{p}(x_0) dx_0\right]\left[\int x_0^2 \overline{p}(x_0) dx_0\right] \widetilde{\psi}_2(y_1) - 2\left[\int q(x \mid x_0) \overline{p}(x_0) dx_0\right]\left[\int x_0 \overline{p}(x_0) dx_0\right]^2 \psi_1^2(y_1).
\end{aligned}
$$

Now,

$$
\begin{aligned}
\int q(x \mid x_0) \overline{p}(x_0) dx_0 &= \overline{p}(x) \\
\int x_0 \overline{p}(x_0) dx_0 &= 0 \\
\int x_0^2 \overline{p}(x_0) dx_0 &= \sigma^2 \\
\int q(x \mid x_0) x_0 \overline{p}(x_0) dx_0 &= \lambda x \overline{p}(x) \\
\int q(x \mid x_0) x_0^2 \overline{p}(x_0) dx_0 &= (\lambda^2 x^2 + 1)\overline{p}(x)
\end{aligned}
$$

This is because

$$
\begin{aligned}
\int q(x \mid x_0) x_0^2 \overline{p}(x_0) dx_0 &= \int x_0 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \lambda x_0)^2}{2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_0^2}{2\sigma^2}\right) dx_0 \\
&= \int \frac{1}{\sqrt{2\pi(1+\lambda^2\sigma^2)}} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1+\lambda^2\sigma^2}}} x_0^2 \exp\left[\frac{\left(x_0 - \lambda\frac{\sigma^2}{1+\lambda^2\sigma^2}x\right)^2}{2\frac{\sigma^2}{1+\lambda^2\sigma^2}}\right] \exp\left[-\frac{x^2}{2(1+\lambda^2\sigma^2)}\right] dx_0 \\
&= \left\{\left[\lambda\frac{\sigma^2}{1+\lambda^2\sigma^2}x\right]^2 + \frac{\sigma^2}{1+\lambda^2\sigma^2}\right\} \frac{1}{\sqrt{2\pi(1+\lambda^2\sigma^2)}} \exp\left[-\frac{x^2}{2(1+\lambda^2\sigma^2)}\right] \\
&= (\lambda^2 x^2 + 1)\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right),
\end{aligned}
$$

where we have used the fact that $1 + \lambda^2\sigma^2 = \sigma^2$. Then,

$$\begin{aligned}
\phi_{\eta\eta}(\overline{p}, y_1, 0)[x] &= (\lambda^2 x^2 + 1)\overline{p}(x)\widetilde{\psi}_2(y_1) - \overline{p}(x)\sigma^2\widetilde{\psi}_2(y_1) \\
&= \overline{p}(x)\widetilde{\psi}_2(y_1)(\lambda^2 x^2 + 1 - \sigma^2) \\
&= \overline{p}(x)\widetilde{\psi}_2(y_1)\lambda^2(x^2 - \sigma^2).
\end{aligned}$$

Combining all of the above, we get:

$$\begin{aligned}
p_1^{[2]}(\eta)[x] &= \overline{p}(x) - \eta\psi_1(y_1)\lambda x\overline{p}(x) + \frac{1}{2}\eta^2\overline{p}(x)\widetilde{\psi}_2(y_1)\lambda^2(x^2 - \sigma^2) \\
&= \overline{p}(x)\left[1 - \eta\psi_1(y_1)\lambda x + \frac{1}{2}\eta^2\widetilde{\psi}_2(y_1)\lambda^2(x^2 - \sigma^2)\right] \\
&= \overline{p}(x)\left[1 - \eta\psi_1(y_1)\lambda x + \frac{1}{2}\eta^2(\psi_2(y_1) + \psi_1(y_1)^2)\lambda^2(x^2 - \sigma^2)\right]
\end{aligned}$$

where

$$\psi_2(y_1) \equiv \widetilde{\psi}_2(y_{t+1}) - \psi_1(y_1)^2 = y_{t+1}\frac{p'_\epsilon(y_{t+1})}{p_\epsilon(y_{t+1})} + y_{t+1}^2\left[\frac{p''_\epsilon(y_{t+1})}{p_\epsilon(y_{t+1})} - \left(\frac{p'_\epsilon(y_{t+1})}{p_\epsilon(y_{t+1})}\right)^2\right].$$

So letting

$$\begin{aligned}
A_{1,0} &= 0, \quad A_{1,1} = \lambda\left[A_{1,0} - \psi_1(y_1)\right] \\
A_{2,0} &= 0, \quad A_{2,1} = \lambda^2\left[A_{2,0} + \psi_2(y_1)\right]
\end{aligned}$$

we have:

$$p_1^{[2]}(\eta)[x] = \overline{p}(x)\left[1 + A_{1,1}\eta x + \frac{1}{2}(A_{2,t} + A_{1,1}^2)\eta^2(x^2 - \sigma^2)\right].$$

Note that $\int |p_1^{[2]}(\eta)[x]|dx < \infty$ so $p_1^{[2]}(\eta) \in L_1(\mathbb{R})$, and $\int p_1^{[2]}(\eta)[x]dx = 1$. Moreover, it follows directly from Lemma 3 that

$$\begin{aligned}
\|p_1(\eta) - p_1^{[2]}(\eta)\|_1 &= o(|\eta|^2) \quad a.s. \\
\int \left|x^k\left(p_1(\eta)[x] - p_1^{[2]}(\eta)[x]\right)\right| dx &= o(|\eta|^2) \quad a.s. \quad \text{for every } k \geqslant 1,
\end{aligned}$$

so that we can call $p_1^{[2]}(\eta)$ a second order approximation to $p_1(\eta)$.

Recursion any t. Start with $p_t^{[2]}(\eta)$ with $p_t^{[2]}(\eta)[x] \equiv \overline{p}(x)\left[1 + A_{1,t}\eta x + 1/2(A_{2,t} + A_{1,t}^2)\eta^2(x^2 - \sigma^2)\right]$, where

$$\begin{aligned}
A_{1,t} &= \lambda\left[A_{1,t-1} - \psi_1(y_t)\right] \\
A_{2,t} &= \lambda^2\left[A_{2,t-1} + \psi_2(y_t)\right],
\end{aligned}$$

and

$$\|p_t(\eta) - p_t^{[2]}(\eta)\|_1 = o(|\eta|^2) \quad a.s.$$

$$\int \left| x^k \Big(p_t(\eta)[x] - p_t^{[2]}(\eta)[x] \Big) \right| dx = o(|\eta|^2) \quad a.s. \quad \text{for every } k \geqslant 1.$$

The goal is to establish that the same property holds at $t+1$. For this, define $\rho_t = p_t(\eta) - p_t^{[2]}(\eta)$ and

$$\begin{aligned}
\widetilde{p}_{t+1}(\eta) &\equiv \phi(p_t(\eta), y_{t+1}, 0) + \phi_\eta(p_t(\eta), y_{t+1}, 0)\eta + \frac{1}{2}\phi_{\eta\eta}(p_t(\eta), y_{t+1}, 0)\eta^2 \\
&= \phi(p_t^{[2]}(\eta) + \rho_t, y_{t+1}, 0) + \phi_\eta(p_t^{[2]}(\eta) + \rho_t, y_{t+1}, 0)\eta + \frac{1}{2}\phi_{\eta\eta}(p_t^{[2]}(\eta) + \rho_t, y_{t+1}, 0)\eta^2.
\end{aligned}$$

As before:

$$\begin{aligned}
\phi(p_t(\eta), y_{t+1}, 0)[x] &= \int q(x \mid x_t) p_t(\eta)[x_t] dx_t, \\
\phi_\eta(p_t(\eta), y_{t+1}, 0)[x] &= -\psi_1(y_{t+1}) \left\{ \int q(x \mid x_t) x_t p_t(\eta)[x_t] dx_t - \left[\int q(x \mid x_t) p_t(\eta)[x_t] dx_t \right] \left[\int x_t p_t(\eta)[x_t] dx_t \right] \right\} \\
\phi_{\eta\eta}(p_t(\eta), y_{t+1}, 0)[x] &= \left[\int q(x \mid x_t) x_t^2 p_t(\eta)[x_t] dx_t \right] \widetilde{\psi}_2(y_{t+1}) \\
&\quad - 2 \left[\int q(x \mid x_t) x_t p_t(\eta)[x_t] dx_t \right] \left[\int x_t p_t(\eta)[x_t] dx_t \right] \psi_1^2(y_{t+1}) \\
&\quad - \left[\int q(x \mid x_t) p_t(\eta)[x_t] dx_t \right] \left[\int x_t^2 p_t(\eta)[x_t] dx_t \right] \widetilde{\psi}_2(y_{t+1}) \\
&\quad - 2 \left[\int q(x \mid x_t) p_t(\eta)[x_t] dx_t \right] \left[\int x_t p_t(\eta)[x_t] dx_t \right]^2 \psi_1^2(y_{t+1}).
\end{aligned}$$

For the constant term, we therefore have:

$$\begin{aligned}
\phi(p_t^{[2]}(\eta) + \rho_t, y_{t+1}, 0)[x] &= \int q(x \mid x_t) \overline{p}(x_t) \left[1 + A_{1,t}\eta x_t + 1/2(A_{2,t} + A_{1,t}^2)\eta^2 (x_t^2 - \sigma^2) \right] dx_t + \int q(x \mid x_t) \rho_t(x_t) dx_t \\
&= \int q(x \mid x_t) \overline{p}(x_t) dx_t + \int q(x \mid x_t) \overline{p}(x_t) A_{1,t}\eta x_t dx_t + \\
&\quad + \frac{1}{2} \int q(x \mid x_t) \overline{p}(x_t) (A_{2,t} + A_{1,t}^2)\eta^2 (x_t^2 - \sigma^2) dx_t + \int q(x \mid x_t) \rho_t(x_t) dx_t \\
&= \overline{p}(x) + A_{1,t}\eta \int q(x \mid x_t) \overline{p}(x_t) x_t dx_t + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2 \int q(x \mid x_t) \overline{p}(x_t)(x_t^2 - \sigma^2) dx_t \\
&\quad + \int q(x \mid x_t) \rho_t(x_t) dx_t \\
&= \overline{p}(x) \left[1 + \lambda A_{1,t}\eta x + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2 (\lambda^2 x^2 + 1 - \sigma^2) \right] + \int q(x \mid x_t) \rho_t(x_t) dx_t \\
&= \overline{p}(x) \left[1 + \lambda A_{1,t}\eta x + \frac{1}{2}\lambda^2 (A_{2,t} + A_{1,t}^2)\eta^2 (x^2 - \sigma^2) \right] + \int q(x \mid x_t) \rho_t(x_t) dx_t
\end{aligned}$$

Write the first order term as:

$$\phi_\eta(p_t^{[2]}(\eta)+\rho_t,y_{t+1},0)[x] = -\psi_1(y_{t+1})\Bigg\{\int q(x\mid x_t)x_t\{p_t^{[2]}(\eta)[x_t]+\rho_t(x_t)\}dx_t$$
$$-\left[\int q(x\mid x_t)\{p_t^{[2]}(\eta)[x_t]+\rho_t(x_t)\}dx_t\right]\left[\int x_t\{p_t^{[2]}(\eta)[x_t]+\rho_t(x_t)\}dx_t\right]\Bigg\}.$$

Now we separately consider the three terms above. First

$$\begin{aligned}
&\int q(x\mid x_t)x_t\{p_t^{[2]}(\eta)[x_t]+\rho_t(x_t)\}dx_t\\
&\qquad=\int q(x\mid x_t)x_t p_t^{[2]}(\eta)[x_t]dx_t+\int q(x\mid x_t)x_t\rho_t(x_t)dx_t\\
&\qquad=\int q(x\mid x_t)\overline{p}(x_t)x_t dx_t + A_{1,t}\eta\int q(x\mid x_t)\overline{p}(x_t)x_t^2dx_t+\\
&\qquad+\frac{1}{2}(A_{2,t}+A_{1,t}^2)\eta^2\int q(x\mid x_t)\overline{p}(x_t)(x_t^3-\sigma^2x_t)dx_t+\int q(x\mid x_t)x_t\rho_t(x_t)dx_t\\
&\qquad=\lambda x\overline{p}(x)+A_{1,t}\eta[1+\lambda^2x^2]\overline{p}(x)+\\
&\qquad+\frac{1}{2}(A_{2,t}+A_{1,t}^2)\eta^2[\lambda^3x^3+3\lambda x-\sigma^2\lambda x_t]\overline{p}(x)+\int q(x\mid x_t)x_t\rho_t(x_t)dx_t.
\end{aligned}$$

Next, as from our calculations above

$$\begin{aligned}
\int q(x\mid x_t)\{p_t^{[2]}(\eta)[x_t]+\rho_t(x_t)\}dx_t &= \overline{p}(x)+\lambda A_{1,t}\eta x\overline{p}(x)+\frac{1}{2}\lambda^2(A_{2,t}+A_{1,t}^2)\eta^2(x^2-\sigma^2)\overline{p}(x)\\
&\quad+\int q(x\mid x_t)\rho_t(x_t)dx_t,
\end{aligned}$$

and finally

$$\begin{aligned}
\int x_t\{p_t^{[2]}(\eta)[x_t]+\rho_t(x_t)\}dx_t &= A_{1,t}\eta\int\overline{p}(x_t)x_t^2dx_t+\frac{1}{2}(A_{2,t}+A_{1,t}^2)\eta^2\int\overline{p}(x_t)(x_t^3-x_t\sigma^2)dx_t+\int x_t\rho_t(x_t)dx_t\\
&= A_{1,t}\eta\sigma^2+\int x_t\rho_t(x_t)dx_t.
\end{aligned}$$

In the above, we have used the facts that:

$$\begin{aligned}
\int q(x\mid x_t)\overline{p}(x_t)x_tdx_t &= \lambda x\overline{p}(x)\\
\int \overline{p}(x_t)x_t^2dx_t &= \sigma^2\\
\int q(x\mid x_t)\overline{p}(x_t)x_t^2dx_t &= \left[1+\lambda^2x^2\right]\overline{p}(x),\\
\int q(x\mid x_t)\overline{p}(x_t)x_t^3dx_t &= \left[3\lambda x+\lambda^3x^3\right]\overline{p}(x).
\end{aligned}$$

Now, combine all of the above to get

$$\phi_\eta(p_t^{[2]}(\eta)+\rho_t,y_{t+1},0)[x] = -\psi_1(y_{t+1})\overline{p}(x)\left\{\lambda x+\lambda^2A_{1,t}\eta(x^2-\sigma^2)\right\}+R_{1,t+1}(x)$$

where

$$
\begin{aligned}
&R_{1,t+1}(x) \equiv \\
&-\psi_1(y_{t+1})\Bigg\{\frac{1}{2}(A_{2,t}+A_{1,t}^2)\eta^2[\lambda^3x^3+3\lambda x-\sigma^2\lambda x_t]\overline{p}(x)+\int q(x\mid x_t)x_t\rho_t(x_t)dx_t \\
&\qquad -\left[\lambda A_{1,t}\eta x\overline{p}(x)+\frac{1}{2}\lambda^2(A_{2,t}+A_{1,t}^2)\eta^2(x^2-\sigma^2)\overline{p}(x)+\int q(x\mid x_t)\rho_t(x_t)dx_t\right][A_{1,t}\eta\sigma^2] \\
&\qquad -\left[\overline{p}(x)+\lambda A_{1,t}\eta x\overline{p}(x)+\frac{1}{2}\lambda^2(A_{2,t}+A_{1,t}^2)\eta^2(x^2-\sigma^2)\overline{p}(x)+\int q(x\mid x_t)\rho_t(x_t)dx_t\right]\left[\int x_t\rho_t(x_t)dx_t\right]\Bigg\}.
\end{aligned}
$$

Using a reasoning similar to that used in the proof of Theorem 1, the result of Lemma 3 implies that

$$
\|R_{1,t+1}\|_1 = O(|\eta|^2) \quad a.s. \tag{38}
$$

Lastly, for the second-order term, write it as:

$$
\phi_{\eta\eta}(p_t^{[2]}(\eta)+\rho_t,y_{t+1},0)[x] = \lambda^2\widetilde{\psi}_2(y_{t+1})\overline{p}(x)(x^2-\sigma^2)+R_{2,t+1}(x),
$$

with

$$
\begin{aligned}
R_{2,t+1}(x) \equiv & \widetilde{\psi}_2(y_{t+1}) \int q(x \mid x_t) x_t^2 \rho_t(x_t) dx_t \\
& -2\psi_1^2(y_{t+1}) \left[\int q(x \mid x_t) x_t p_t^{[2]}(\eta)[x_t] dx_t \right] \left[\int x_t \rho_t(x_t) dx_t \right] \\
& -2\psi_1^2(y_{t+1}) \left[\int q(x \mid x_t) x_t \rho_t(x_t) dx_t \right] \left[\int x_t (p_t^{[2]}(\eta)[x_t] + \rho_t(x_t)) dx_t \right] \\
& -\widetilde{\psi}_2(y_{t+1}) \left[\int q(x \mid x_t) p_t^{[2]}(\eta)[x_t] dx_t \right] \left[\int x_t^2 \rho_t(x_t) dx_t \right] \\
& -\widetilde{\psi}_2(y_{t+1}) \left[\int q(x \mid x_t) \rho_t(x_t) dx_t \right] \left[\int x_t^2 (p_t^{[2]}(\eta)[x_t] + \rho_t(x_t)) dx_t \right] \\
& -2\psi_1^2(y_{t+1}) \left[\int q(x \mid x_t) p_t^{[2]}(\eta)[x_t] dx_t \right] \left[\int x_t \rho_t(x_t) dx_t \right]^2 \\
& -2\psi_1^2(y_{t+1}) \left[\int q(x \mid x_t) \rho_t(x_t) dx_t \right] \left[\int x_t (p_t^{[2]}(\eta)[x_t] + \rho_t(x_t)) dx_t \right]^2 \\
& -4\psi_1^2(y_{t+1}) \left[\int q(x \mid x_t) p_t^{[2]}(\eta)[x_t] dx_t \right] \left[\int x_t \rho_t(x_t) dx_t \right] \left[\int x_t p_t^{[2]}(\eta)[x_t] dx_t \right] \\
& +\widetilde{\psi}_2(y_{t+1}) \overline{p}(x) \left\{ A_{1,t}\eta(\lambda^3 x^3 + 3\lambda x) + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^4 x^4 + 6\lambda^2 x^2 + 3 - \sigma^2(\lambda^2 x^2 + 1)) \right\} \\
& -2\psi_1^2(y_{t+1}) A_{1,t}\eta\sigma^2 \overline{p}(x) \left\{ A_{1,t}\eta(\lambda^2 x^2 + 1) + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^3 x^3 + 3\lambda x - \sigma^2 \lambda x) \right\} \\
& -\widetilde{\psi}_2(y_{t+1}) \overline{p}(x) \left\{ 1 + A_{1,t}\eta\lambda x_t + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^2 x^2 + 1 - \sigma^2) \right\} \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2 2\sigma^4 \\
& -\widetilde{\psi}_2(y_{t+1}) \overline{p}(x) \sigma^2 \left\{ A_{1,t}\eta\lambda x + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^2 x^2 + 1 - \sigma^2) \right\} \\
& -2\psi_1^2(y_{t+1}) \left(A_{1,t}\eta\sigma^2 \right)^2 \overline{p}(x) \left\{ 1 + A_{1,t}\eta\lambda x + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^2 x^2 + 1 - \sigma^2) \right\},
\end{aligned}
$$

where we used the following results:

$$
\begin{aligned}
& \int q(x \mid x_t) x_t^2 p_t^{[2]}(\eta)[x_t] dx_t \\
& = \int q(x \mid x_t) \overline{p}(x_t) \left[x_t^2 + A_{1,t}\eta x_t^3 + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(x_t^4 - \sigma^2 x_t^2) \right] dx_t \\
& = \overline{p}(x) \left\{ \lambda^2 x^2 + 1 + A_{1,t}\eta(\lambda^3 x^3 + 3\lambda x) + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^4 x^4 + 6\lambda^2 x^2 + 3 - \sigma^2(\lambda^2 x^2 + 1)) \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
\int q(x \mid x_t) x_t p_t^{[2]}(\eta)[x_t] dx_t &= \int q(x \mid x_t)\overline{p}(x_t)\left[x_t + A_{1,t}\eta x_t^2 + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(x_t^3 - \sigma^2 x_t)\right] dx_t \\
&= \overline{p}(x)\left\{\lambda x + A_{1,t}\eta(\lambda^2 x^2 + 1) + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^3 x^3 + 3\lambda x - \sigma^2 \lambda x)\right\} \\
\int q(x \mid x_t) p_t^{[2]}(\eta)[x_t] dx_t &= \int q(x \mid x_t)\overline{p}(x_t)\left[1 + A_{1,t}\eta x_t + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(x_t^2 - \sigma^2)\right] dx_t \\
&= \overline{p}(x)\left\{1 + A_{1,t}\eta\lambda x + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(\lambda^2 x^2 + 1 - \sigma^2)\right\} \\
\int x_t^2 p_t^{[2]}(\eta)[x_t] dx_t &= \int \overline{p}(x_t)\left[x_t^2 + A_{1,t}\eta x_t^3 + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(x_t^4 - \sigma^2 x_t^2)\right] dx_t \\
&= \sigma^2 + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2 2\sigma^4 \\
\int x_t p_t^{[2]}(\eta)[x_t] dx_t &= \int \overline{p}(x_t)\left[x_t + A_{1,t}\eta x_t^2 + \frac{1}{2}(A_{2,t} + A_{1,t}^2)\eta^2(x_t^3 - \sigma^2 x_t)\right] dx_t \\
&= A_{1,t}\eta\sigma^2.
\end{aligned}
$$

Note that similar to before, the result of Lemma 3 implies that

$$
\|R_{2,t+1}\|_1 = O(|\eta|) \quad a.s. \tag{39}
$$

Putting everything together, and letting $\widetilde{R}_{1,t+1}(x) \equiv \int q(x \mid x_t)\rho_t(x_t)dx_t + \eta R_{1,t+1}(x) + \eta^2 R_{2,t+1}(x)$, then gives

$$
\begin{aligned}
\widetilde{p}_{t+1}(\eta)[x] =& \overline{p}(x)\left[1 + \lambda A_{1,t}\eta x + \frac{1}{2}\lambda^2(A_{2,t} + A_{1,t}^2)\eta^2(x^2 - \sigma^2)\right] + \int q(x \mid x_t)\rho_t(x_t)dx_t \\
&- \psi_1(y_{t+1})\overline{p}(x)\left\{\lambda x\eta + \lambda^2 A_{1,t}\eta^2(x^2 - \sigma^2)\right\} + \eta R_{1,t+1}(x) \\
&+ \frac{1}{2}\lambda^2\widetilde{\psi}_2(y_{t+1})\overline{p}(x)(x^2 - \sigma^2)\eta^2 + R_{2,t+1}(x)\eta^2 \\
&= \overline{p}(x)\left\{1 + \eta x\lambda\left(A_{1,t} - \psi_1(y_{t+1})\right) + \frac{1}{2}\eta^2\lambda^2\left(A_{2,t} + A_{1,t}^2 - 2\psi_1(y_{t+1})A_{1,t} + \widetilde{\psi}_2(y_{t+1})\right)(x^2 - \sigma^2)\right\} \\
&+ \widetilde{R}_{t+1}(x) \\
&= \overline{p}(x)\left\{1 + \eta x A_{1,t+1} + \frac{1}{2}\eta^2\left(A_{2,t+1} + A_{1,t+1}^2\right)(x^2 - \sigma^2)\right\} + \widetilde{R}_{t+1}(x) \\
&= p_{t+1}^{[2]}(\eta)[x] + \widetilde{R}_{t+1}(x),
\end{aligned}
$$

where we have let

$$
\begin{aligned}
A_{1,t+1} &= \lambda\left(A_{1,t} - \psi_1(y_{t+1})\right) \\
A_{2,t+1} &= \lambda^2(A_{2,t} + \widetilde{\psi}_2(y_{t+1}) - \psi_1(y_{t+1})^2)
\end{aligned}
$$

so that

$$\begin{aligned} A_{2,t+1} + A_{1,t+1}^2 &= \lambda^2(A_{2,t} + \widetilde{\psi}_2(y_{t+1}) - \psi_1(y_{t+1})^2) + (\lambda(A_{1,t} - \psi_1(y_{t+1})))^2 \\ &= A_{2,t} + A_{1,t}^2 - 2\psi_1(y_{t+1})A_{1,t} + \widetilde{\psi}_2(y_{t+1}). \end{aligned}$$

Triangle inequality combined with (38) and (39) yields $\|\widetilde{R}_{t+1}\|_1 = o(|\eta|^2)$ a.s. A reasoning similar to that in the proof of Theorem 1 combined with Lemma 3 also ensures that $\|(\cdot)^k \widetilde{R}_{t+1}(\cdot)\|_1 = o(|\eta|^2)$ a.s. Therefore the property holds at $t+1$.

B.2 Third Order Approximation

We state the expression of the third order approximation without detailed derivations which are straightforward though tedious. The third-order approximation is given by:

$$p_t^{[3]}(\eta)[x] \equiv \overline{p}(x)\left[1 + \sum_{m=1}^{3} \frac{1}{m!}\eta^m \left(\sum_{j=0}^{m} A_{m,j,t}x^j\right)\right],$$

where the constants $A_{m,j,t}$ are determined as follows. For the linear term, $A_{1,0,t} = 0$ and

$$A_{1,1,t} \equiv A_{1,t} = \lambda\left[A_{1,t-1} - \psi_1(y_t)\right], \quad t \geqslant 1, \quad A_{1,0} = 0$$

where $\psi_1(y) = 1 + yp'_\epsilon(y)/p_\epsilon(y)$. For the second-order term, $A_{2,0,t} = -A_{2,2,t}(1-\lambda^2)^{-1}$, $A_{2,1,t} = 0$, $A_{2,2,t} = A_{2,t} - A_{1,t}^2$, where

$$A_{2,t} = \lambda^2(A_{2,t-1} + \psi_2(y))$$

and $\psi_2(y) = y\frac{p'_\epsilon(y)}{p_\epsilon(y)} + y^2\left(\frac{p''_\epsilon(y)}{p_\epsilon(y)} - \left(\frac{p'_\epsilon(y)}{p_\epsilon(y)}\right)^2\right)$. For the third-order term, $A_{3,0,t} = 0$ and $A_{3,2,t} = 0$ and

$$\begin{aligned} A_{3,3,t} &= \lambda^3(A_{3,3,t-1} + \psi_{3,3}(y_t) + 3\psi_{3,2}(y_t)A_{1,t-1} + 3\psi_{3,1}(y_t)A_{2,2,t-1}), \\ A_{3,1,t} &= 3\lambda^{-2}A_{3,3,t} + \lambda\Big[A_{3,1,t-1} - 3(1-\lambda^2)^{-1}\psi_{3,1}(y_t)A_{2,2,t} \\ &- 3(1-\lambda^2)^{-1}\left(\psi_{3,2}(y_t) + 2\psi_{3,1}(y_t)A_{1,t-1}\right)\left(\psi_{3,1}(y_t) + A_{1,t-1}\right)\Big], \end{aligned}$$

where

$$\begin{aligned} \psi_{3,3}(y) &= \frac{1}{x^3 p(y|x,\eta)} \left.\frac{\partial^3 p(y|x,\eta)}{\partial \eta^3}\right|_{\eta=0} = -1 - 7\frac{p'_\epsilon(y)}{p_\epsilon(y)}y - 6\frac{p''_\epsilon(y)}{p_\epsilon(y)}y^2 - \frac{p'''_\epsilon(y)}{p_\epsilon(y)}y^3, \\ \psi_{3,2}(y) &= \frac{1}{x^2 p(y|x,\eta)} \left.\frac{\partial^2 p(y|x,\eta)}{\partial \eta^2}\right|_{\eta=0} = 1 + 3\frac{p'_\epsilon(y)}{p_\epsilon(y)}y + \frac{p''_\epsilon(y)}{p_\epsilon(y)}y^2, \\ \psi_{3,1}(y) &= \frac{1}{x p(y|x,\eta)} \left.\frac{\partial p(y|x,\eta)}{\partial \eta}\right|_{\eta=0} = -1 - \frac{p'_\epsilon(y)}{p_\epsilon(y)}y. \end{aligned}$$

References

Andersen, T. G., L. Benzoni, and J. Lund (2002): "An Empirical Investigation of Continuous-Time Equity Return Models," *The Journal of Finance*, 57(3), 1239–1284.

Anderson, B., and J. Moore (1979): *Optimal Filtering*. Prentice-Hall, Inc.

Baker, Jr., G., and P. Graves-Morris (1996): *Padé Approximants*. Cambridge University Press, London.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59(4), 1481–1509.

Bensoussan, A. (1988): *Perturbation methods in optimal control*. John Wiley & Sons.

Bertsekas, D. P., and S. E. Shreve (1978): *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press.

Bollerslev, T. (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69(3), 542 – 547.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22(11), 4463–4492.

Caldara, D., J. Fernández-Villaverde, P. A. Guerrón-Quintana, J. Rubio-Ramírez, and W. Yao (2011): "Computing DSGE Models with Recursive Preferences and Stochastic Volatility," *Review of Economic Dynamics*, 15(2), pp. 188 – 206.

Creal, D. (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 31(3), 245 – 296.

Crisan, D., and A. Doucet (2002): "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, 50(3), 736–746.

Douc, R., G. Fort, E. Moulines, and P. Priouret (2009): "Forgetting the initial distribution for Hidden Markov Models," *Stochastic Processes and their Applications*, 119(4), 1235 – 1256.

Fernández-Villaverde, J., P. Guerrón-Quintana, and J. F. Rubio-Ramírez (2015): "Estimating dynamic equilibrium models with stochastic volatility," *Journal of Econometrics*, 185(1), 216 – 229.

Fernández-Villaverde, J., P. A. Guerrón-Quintana, J. Rubio-Ramírez, and M. Uribe (2011): "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, 101, pp. 2530 – 2561.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2006): "Solving DSGE models with perturbation methods and a change of variables," *Journal of Economic Dynamics and Control*, 30(12), 2509 – 2531.

——— (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *The Review of Economic Studies*, 74(4), 1059 – 1087.

Fleming, W. (1971): "Stochastic Control for Small Noise Intensities," *SIAM Journal on Control*, 9(3), 473–517.

Ghysels, E., A. Harvey, and E. Renault (1996): *Handbook of Statistics*vol. 14, chap. Stochastic volatility. North-Holland, Amsterdam.

Hansen, L. P., and T. J. Sargent (2007): "Recursive robust estimation and control without commitment," *Journal of Economic Theory*, 136(1), 1 – 27.

——— (2010): "Fragile beliefs and the price of uncertainty," *Quantitative Economics*, 1(1), 129 – 162.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *The Review of Economic Studies*, 61(2), 247 – 264.

Hull, J., and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *The Journal of Finance*, 42(2), 281–300.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12(4), 371 – 389.

Judd, K. L. (1996): *Oxford Handbook of Innovation*vol. I, chap. Approximation, Perturbation, and Projection Methods in Economic Analysis. Handbook of Computational Economics.

——— (2002): "Perturbation Methods with Nonlinear Changes of Variables," Working paper, Hoover Istitution, Stanford University.

Justiniano, A., and G. E. Primicieri (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), pp. 604–641.

Kim, J., S. Kim, E. Schaumburg, and C. A. Sims (2008): "Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models," *Journal of Economic Dynamics & Control*, 32, 3397 – 3414.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH," *The Review of Economic Studies*, 65(3), 361 – 393.

Kizner, W. (1969): "Optimal Nonlinear Estimation Based on Orthogonal Expansions," Technical Report 1332 - 1366, Jet Propulsion Laboratory, California Institute of Technology.

Meyn, S., and R. Tweedie (1993): *Markov Chains and Stochastic Stability.* Springer, London.

Polson, N. G., J. R. Stroud, and P. Müller (2008): "Practical Filtering with Sequential Parameter Learning," *Journal of the Royal Statistical Society*, 70(2), 413 – 428.

Ruiz, E. (1994): "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289 – 306.

Schmitt-Grohé, S., and M. Uribe (2004): "Solving dynamic general equilibrium models using a second-order approximation to the policy function," *Journal of Economic Dynamics and Control*, 28(4), 755 – 775.

Shephard, N. (2004): *Stochastic Volatility: Selected Readings*. Oxford University Press, Oxford.

Shephard, N., and M. Pitt (1997): "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84(3), 653 – 667.

——— (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446), 590 – 599.

Sorenson, H. W., and A. R. Stubberud (1968): "Non-linear filtering by approximation of the posterior density," *International Journal of Control*, 8(1), 33 – 51.