Robust Dynamic Estimation

Olivier Rubel,[a] Prasad A. Naik[a]

[a] Graduate School of Management, University of California Davis, Davis, California 95616
Contact: orubel@ucdavis.edu (OR); panaik@ucdavis.edu (PAN)

Abstract. Managing marketing resources over time requires dynamic model estimation, which necessitates specifying some parametric or nonparametric probability distribution. When the data generating process differs from the assumed distribution, the resulting model is misspecified. To hedge against such a misspecification risk, the extant theory recommends using the sandwich estimator. This approach, however, only corrects the *variance* of estimated parameters, but not their *values*. Consequently, the sandwich estimator does not affect any managerial outcomes such as marketing budgeting and allocation decisions. To overcome this drawback, we present the minimax framework that does not necessitate distributional assumptions to estimate dynamic models. Applying minimax control theory, we derive an optimal robust filter, illustrate its application to a unique advertising data set from the Canadian Blood Services, and contribute several novel findings. We discover the compensatory effect: Advertising effectiveness increases and the carryover effect decreases as robustness increases. We also find that the robust filter uniformly outperforms the Kalman filter on the out-of-sample predictions. Furthermore, we uncover the existence of a profit-volatility trade-off, similar to the returns-risk trade-off in finance, whereby the volatility of profit stream decreases at the expense of reduced total profit as robustness increases. Finally, we prove that, unlike for-profit companies, managers of nonprofit organizations should optimally allocate budgets *opposite* the advertising-to-sales ratio heuristic; that is, advertise more (less) when sales are low (high).

1. Introduction

Brand managers should optimally allocate marketing resources to maximize long-term profit (Gatignon 1993, Mantrala 2002, Shankar 2008, Gupta and Steenburgh 2008). To this end, they seek counsel from marketing analytics companies (e.g., IRI, MarketShare, Kvantum), who analyze the firm's own market data augmented with data from other sources (e.g., Kantar, Nielsen, J.D. Power, etc.), assess the impact of the marketing activities on sales performance, and then recommend optimal budgets and allocations for brand managers to consider when developing marketing plans. For real-world case studies documenting this disciplined practice, see Corstjens and Merrihue (2003), Kumar and Mirchandani (2012), or Nichols (2013). To furnish optimal recommendations, marketing analytics companies use models and methods often invented by academic researchers from various scientific disciplines. For instance, marketing mix models find their roots in the pioneering work of Little (1975), while methods for estimating dynamic models rest on the foundations laid by Fisher's (1922) likelihood principle and Kalman's (1960) filter. In summary, brand managers make the decisions on marketing budgets and allocations; marketing analytics companies conduct the analyses using existing models and methods to facilitate marketing decisions; and academic researchers invent new models and methods that enable informed decision making and thus improve the practice of marketing.

The optimal allocation of marketing resources over time requires estimation of dynamic models that relate marketing activities (e.g., advertising, price, etc.) to market outcomes (e.g., sales, share, etc.). Existing methods to estimate such dynamic models assume that the forecast errors between the model's predictions and the market data follow *some* probability distribution, which is typically multivariate normal (e.g., Naik et al. 1998, Vakratsas et al. 2004, Lachaab et al. 2006, Ataman et al. 2008, Bruce 2008, Rubel et al. 2011, van Heerde et al. 2013, Liu and Shankar 2015, Kolsarici and Vakratsas 2015). The data generating process can, however, differ from the assumed probability distribution. The resulting estimated parameters depend on this ex ante assumption of some distribution F_i. Yet what if the distribution was F_j instead? Furthermore,

the probability distributions are infinite and can be represented in the set $\Omega = \{F_1, \ldots, F_i, \ldots, F_\infty\}$, where F_i can be Normal, Gamma, Gaussian Mixture, etc. Because there exists an infinite number of distributions in Ω, it is unreasonable to expect researchers or analysts to estimate parameters for each of the alternative distributions. Hence we propose to develop a new method for estimating dynamic models that does not require the assumption or specification of any probability distribution, parametric or nonparametric.

To elucidate the conceptual idea, let ω_{Normal} denote the forecast errors resulting under the assumption of a Normal distribution; ω_{Gamma} under the Gamma distribution; and so on. Then the set $\omega = \{\omega_1, \ldots, \omega_i, \ldots, \omega_\infty\}$ contains the forecast errors arising from the corresponding distributions in Ω. Because the proposed method does not assume any specific distribution, the forecast errors arising from all of the distributions must be accounted for when estimating model parameters. To accommodate all of the distributions, we consider the largest possible errors $\omega^* = \max\{\omega_{\text{Normal}}, \omega_{\text{Gamma}}, \omega_{\text{Gaussian Mixture}}, \ldots, \omega_i, \ldots, \omega_\infty\}$. Consequently, regardless of any specific distribution, we conservatively estimate parameters under the worst case scenario stemming from the largest possible number of errors. As a result, the estimated parameters are *robust* to any distribution in the set Ω of all possible distributions. To summarize, robustness is a conceptual property pertaining to the possibility set Ω, while conservatism operationalizes robustness via the worst case scenario.

We formalize the notion of worst case scenarios via the minimax principle (Wald 1950). To understand this principle, let $L(\alpha, \omega)$ denote a loss function such as the sum of squared errors, where α denotes the model parameters and ω are the forecast errors. In the extant literature, one assumes that ω follows some probability distribution (typically Normal), then evaluates the expected loss $L(\alpha) = \mathbb{E}[L(\alpha, \omega)]$, and finally minimizes it with respect to α to obtain the parameter estimates $\hat{\alpha}$. By contrast, the proposed method eschews the distributional assumption on ω, minimizes the loss function $L(\alpha, \omega)$ with respect to α *and* maximizes it with respect to ω, and thus obtains the parameter estimates $\hat{\alpha}$. Formally, we solve the minimax problem

$$\underbrace{\min}_{\alpha} \; \underbrace{\max}_{\omega} \; L(\alpha, \omega)$$

to obtain the (best) parameter estimates $\hat{\alpha}$ under the largest errors ω^* (worst case). In the context of dynamic models, this minimax problem transcends to a multi-period game. Applying minimax control theory (see, e.g., Başar and Bernhard 1995), we solve this dynamic game to obtain the best parameter estimates $\{\hat{\alpha}_t: t = 1, \ldots, N\}$ under the largest possible errors $\{\omega_t^*: t = 1, \ldots, N\}$.

We present the results in Propositions 1 and 2. Specifically, Proposition 1 furnishes ω_t^* for every period t without making *any* distributional assumptions. Proposition 2 offers $\hat{\alpha}_t$ recursively over time t, and depends on a scalar parameter γ, which specifies the degree of conservatism (or robustness): The smaller the γ, the more conservative (or robust) the estimation. Interestingly, we show that the robust recursions of $\hat{\alpha}_t$ converge to the standard Kalman filter recursions as γ tends to infinity. In other words, we show that the Kalman filter estimation is the *least* conservative. As γ decreases, the proposed robust filter becomes more conservative, which means it places more weight on the observed data than the Kalman filter, thereby accommodating a broader range of scenarios.

In summary, the proposed robust filter not only relaxes the necessity of specifying distributions of any kind but also facilitates the robust estimation of time-varying dynamic linear models, which cover broad classes such as the vector auto regression models (e.g., Lütkepohl 2005), ARIMA models (e.g., Brockwell and Davis 1996), time-varying parameter models (e.g., Koop and Korobilis 2010), dynamic regression (e.g., Biyalogorsky and Naik 2003), dynamic factor models (e.g., Bruce et al. 2012) or hierarchical linear models (e.g., Raudenbush and Bryk 2002), to mention a few examples. Thus, for such a broad class of dynamic models, the proposed approach strives to improve the practice of marketing by circumventing the risk of model misspecification. When models are misspecified, the forecast errors contain information, which the proposed robust filter incorporates to learn about the true state by placing more weight on the observed data than the normal Kalman filter.

We apply the proposed dynamic robust estimation to a unique data set from the Canadian Blood Services (CBS), which presents three years of weekly advertising and blood collection across all five provinces of English-speaking Canada (i.e., without Québec). Our analysis reveals that, as conservatism increases, the effectiveness of advertising increases, whereas the carryover effect decreases. This compensatory effect generalizes across all of the provinces. This result arises because the proposed filter places greater weight on more recent data. Furthermore, we compare the dynamic robust estimation with the Kalman filter estimation, with and without White's (1980) correction for robust inferences via the sandwich estimator. We also provide diagnostic tests to determine when to use which method, i.e., robust filter, Kalman filter, robust filter with White's correction, or Kalman filter with White's correction, so that researchers and analysts can use the appropriate method in their empirical contexts. Finally, we derive the optimal budget and allocation strategies for nonprofit organizations whose objective differs from profit-maximization. We prove that, unlike

for-profit companies, they should spend the marketing budget *opposite* the advertising-to-sales ratio heuristic, i.e., advertise more (less) when sales are low (high).

The rest of the paper proceeds as follows. Section 3 describes the framework for robust estimation; Section 4 derives the robust filter; Section 5 applies the robust filter to blood donation data and provides an empirical comparison with the Kalman filter and White's correction; Section 6 furnishes the optimal budget and allocation strategies. Section 7 presents concluding comments. We begin with a review of the extant literature on model misspecification.

2. Literature Review

Model misspecification arises when the actual data generating process differs from the assumed probability distribution. Starting with Eicker (1963), this issue has been studied in statistics (see, e.g., Eicker 1967, Huber 1967, Hinkley 1977) and led to an approach known as heteroskedasticity-consistent estimation (White 1980, 1982). In this approach, the standard errors of the estimated parameters are "corrected" to hedge misspecification risk by modifying the inference conducted via the likelihood theory. For misspecified models, White (1980) shows that the inverse of the Fisher information matrix does not yield a consistent variance-covariance matrix, thus resulting in incorrect statistical inferences. To conduct robust inferences, White (1980) proposes the sandwich estimator, which uses not only the inverse of the Fisher information matrix but also the gradients of the log-likelihood function; Equation (22) in Section 4 presents the formula for computational purposes.

In the context of dynamic models, we estimate parameters by computing the likelihood function using the Kalman filter recursions, and conduct statistical inferences, i.e., significance of the estimated parameters, with or without the sandwich estimator. For example, see Naik et al. (1998) for the Kalman filter estimation without White's correction, and Naik et al. (2008) for robust inferences via the sandwich estimator. Panel A in Figure 1 presents the Kalman filter estimation without White's correction; Panel B in Figure 1 augments it with White's correction for robust inferences.

The primary drawback of White's correction (i.e., the sandwich estimator) is that it does not correct the *values* of the estimated parameters; it only corrects for their *variance*. Consequently, managerial decisions would not change whether one uses the sandwich estimator. Consider, for example, estimating a dynamic sales response model with and without White's correction. Would the estimated level of optimal budget or the optimal allocation to marketing activities change? No, because these quantities are functions of the model parameters, whose estimated values remain the same; only their standard errors change due to White's correction. Hence the recommended budget and its allocation to marketing activities remain unaltered.

To overcome this drawback, we develop a method to "correct" the parameter estimator itself. In doing so, the proposed method also corrects for the variance of the estimated parameters. Recall that Panels A and B in Figure 1 require prespecifying a probability distribution, which we aim to relax. To this end, in Panel C of Figure 1, the proposed method, which we call the *robust filter*, makes no distributional assumptions for the error terms in dynamic linear models.

Because the robust filter provides different parameter estimates than those obtained from Panels A and B

Figure 1. Robust Estimation and Inferences

	Robust inferences?	
	No	**Yes**
No	**Panel A** **Kalman filter** ↓ 1. Nonrobust parameter estimates 2. Nonrobust inferences 3. Nonrobust budget 4. Nonrobust allocation	**Panel B** **White's correction** ↓ 1. Nonrobust parameter estimates 2. Robust inferences 3. Nonrobust budget 4. Nonrobust allocation
Yes	**Panel C** **Robust filter** ↓ 1. Robust parameter estimates 2. Robust inferences 3. Conservative budget 4. Conservative allocation	**Panel D** **Combined approach** ↓ 1. Robust parameter estimates 2. Double robust inferences 3. Conservative budget 4. Conservative allocation

(Robust estimation? — row label, left axis)

456

in Figure 1, managerial outcomes such as the budget and allocation change as do their standard errors. In other words, the robust filter yields the robust estimates and robust inferences. Finally, by applying the sandwich estimator together with the dynamic robust estimation, we achieve "double" robustification for inferences, as shown in Panel D of Figure 1.

In summary, the extant literature investigates robustification through the lens of statistical outcomes such as inference, but it cannot affect any managerial outcomes; managerial decisions do not change with or without the sandwich estimator. How else should managers then safeguard against misspecified models? To address this question, we next present the minimax framework to robustify parameter estimation.

3. A Framework for Robust Dynamic Estimation

Consider the dynamic linear model (e.g., Harvey 1989, West and Harrison 1997, Shumway and Stoffer 2011) given by the observation equation

$$y_t = Z_t \alpha_t + v_t, \tag{1}$$

where y_t is the $n \times 1$ vector of observed dependent variables, α_t is the $m \times 1$ vector of unobserved state variables, Z_t is the $n \times m$ link matrix, and v_t is the $n \times 1$ vector of error terms that follow a multivariate normal distribution $N(0, H_t)$. Furthermore, the state vector α_t dynamically evolves over time via the transition equation

$$\alpha_{t+1} = T_t \alpha_t + d_t + \omega_t, \tag{2}$$

where T_t is the $m \times m$ transition matrix, d_t is the $m \times 1$ drift vector, and the error terms ω_t follow an m-dimensional multivariate normal distribution $N(0, Q_t)$. Because the system matrices $\{Z_t, T_t, d_t, H_t, Q_t\}$ can be time-varying, Equations (1) and (2) encapsulate many classes such as the vector auto regression models (e.g., Lütkepohl 2005), ARIMA models (e.g., Brockwell and Davis 1996), time-varying parameter models (e.g., Koop and Korobilis 2010), dynamic regression (e.g., Biyalogorsky and Naik 2003), dynamic factor models (e.g., Bruce et al. 2012) or hierarchical linear models (e.g., Raudenbush and Bryk 2002), to mention a few examples. Section D in the web appendix explains how these models are special cases of Equations (1) and (2).

In Equations (1) and (2), the state vector α_t is unobserved; hence, we seek the estimator $\hat{\alpha}_t$ of the unobserved state α_t by minimizing $\sum_{t=1}^{N} \|\alpha_t - \hat{\alpha}_t\|^2$ assuming that the error terms (v_t, ω_t) follow multivariate normal distributions. In other words, we minimize the expected sum of squared errors, $L(\alpha) = \mathbb{E}[\sum_{t=1}^{N} \|\alpha_t - \hat{\alpha}_t\|^2]$, to find the best estimator $\hat{\alpha} = \{\hat{\alpha}_t: t = 1, \ldots, N\}$, which is given by the Kalman filter recursions (see, e.g., Appendix B in Naik et al. 1998).

Here we have used the notation $\|x\|^2 = x'x$ and, more generally, $\|x\|_A^2 = x'Ax$ for any vector x and a conformable positive definite matrix A. The loss function $L(\alpha)$ can be viewed as the *expected* loss $L(\alpha) = \mathbb{E}[L(\alpha, v, \omega)]$, where the expectation operator $\mathbb{E}[\cdot]$ integrates out (v, ω) using the assumed distributions of the error terms. This operation explains why statistical estimation requires the assumption of *some* probability distribution for (v, ω).

To not assume any probability distributions for (v, ω), we invoke the minimax principle, due to Wald (1950), and replace (v, ω) by their worst case (i.e., largest) disturbances. Specifically, we seek the estimator $\hat{\alpha} = \{\hat{\alpha}_t: t = 1, \ldots, N\}$ by minimizing $\sum_{t=1}^{N} \|\alpha_t - \hat{\alpha}_t\|^2$ with respect to $\hat{\alpha}_t$, as before, but subject to the *maximization* of the sum of squared disturbances,

$$D(v, \omega) = \sum_{t=0}^{N-1} \left(\|\omega_t\|_{Q_t^{-1}}^2 + \|v_t\|_{H_t^{-1}}^2 \right) + \|\alpha_0 - \hat{\alpha}_0\|_{P_0^{-1}}^2,$$

where $(\alpha_0, P_0, \hat{\alpha}_0)$ denote the initial state, its uncertainty, and its estimate, respectively. In other words, we replace the integration operation $\mathbb{E}[\cdot]$ by the maximization operation so as to eschew the distributional assumptions. Thus, in the robust estimation framework, we make no distributional assumptions for the error terms because the minimax principle optimizes over the unobserved states and the unknown errors.

To construct a joint objective function, we combine the loss functions $L(\alpha)$ and $D(v, \omega)$ in the performance index

$$
\begin{aligned}
&L(\alpha_t, v_t, \omega_t) \\
&= \frac{\sum_{t=1}^{N} \|\alpha_t - \hat{\alpha}_t\|^2}{\|\alpha_0 - \hat{\alpha}_0\|_{P_0^{-1}}^2 + \sum_{t=0}^{N-1} (\|\omega_t\|_{Q_t^{-1}}^2 + \|v_t\|_{H_t^{-1}}^2)}.
\end{aligned} \tag{3}
$$

For further details on this performance index, see Banavar (1992), Theodor and Shaked (1994), Başar and Bernhard (1995) or Simon (2006). In Equation (3), all individual terms on the right-hand side (RHS) are quadratic forms and hence positive, so zero is the least value that $L(\cdot)$ attains. We search for all of the finite sequences of $\{(\hat{\alpha}_t, \omega_t): t = 1, \ldots, N\}$ such that the performance index is smaller than some positive constant γ, i.e., the feasible solutions satisfy $L(\alpha_t, v_t, \omega_t) < \gamma$.

The parameter γ embodies conservatism in designing the optimal state estimator $\hat{\alpha}$. To understand this idea, consider the scenario when γ is small. Then, ceteris paribus, the optimal ω^* must be sufficiently large to satisfy $L(\alpha_t, v_t, \omega_t) < \gamma$. The large ω^* in turn perturbs the state α_t substantially due to Equation (2). By contrast, when γ is large, even small values of ω^* can satisfy $L(\alpha_t, v_t, \omega_t) < \gamma$. The small ω^* in turn marginally perturbs the state α_t due to Equation (2). In the extreme, when γ tends to infinity, the disturbances ω^*

are the smallest, leading to the least conservative scenario. Hence, as γ decreases, the estimation becomes more conservative (or robust) by accommodating wide perturbations to the unobserved state α_t.

Given any level of conservatism γ, what are the best trajectories of the state estimator $\hat{\alpha} = \{\hat{\alpha}_t: t = 1, \ldots, N\}$ and the disturbance sequences $(v^*, \omega^*) = \{(v_t^*, \omega_t^*): t = 0, \ldots, N-1\}$? To find them, we need to solve the minimax problem

$$(\hat{\alpha}_t, v_t^*, \omega_t^*) = \arg \underbrace{\min}_{\alpha_t} \underbrace{\max}_{(v_t, \omega_t)} L(\alpha_t, v_t, \omega_t) < \gamma, \qquad (4)$$

subject to Equations (1) and (2) and without distributional assumptions on (v_t, ω_t).

Equation (4) induces a dynamic game, "dynamic" because Equation (2) evolves over time, and "game" because one can evoke the metaphor of two fictitious players, the so-called statistician and Nature, engaged in a competitive game in which the statistician seeks to estimate the unobserved state α_t, while Nature injects disturbances (v_t, ω_t) to obscure the state and thus impede the statistician. Formally, the statistician chooses $\hat{\alpha}_t \in (-\infty, \infty)$ to minimize $L(\alpha_t, v_t, \omega_t)$, and Nature chooses the disturbances $(v_t^*, \omega_t^*) \in (-\infty, \infty)$ to maximize $L(\alpha_t, v_t, \omega_t)$. Because ω_t^* perturbs the future evolution of α_t (via Equation (2)), the statistician foresees this perturbed path and incorporates its effect in determining the best estimator $\{\hat{\alpha}_t\}$. Likewise, Nature foresees the statistician's best response over time $\{\hat{\alpha}_t\}$ when determining her best response $\{(v_t^*, \omega_t^*)\}$ such that neither of them have incentives to deviate, thereby attaining the equilibrium strategies $\{(\hat{\alpha}_t, v_t^*, \omega_t^*): t = 1, \ldots, N\}$.

In summary, the minimax estimation framework reveals three insights, i.e., no distributional assumptions are necessary to estimate the unknown state α_t in dynamic linear models; the parameter γ embodies conservatism (small $\gamma \Rightarrow$ more conservative); the optimal state estimator $\hat{\alpha}_t$ is the *least* conservative when $\gamma \to \infty$.

4. Deriving the Robust Filter

Here we solve the dynamic game in Equation (4) to derive the method for robust dynamic estimation. We first obtain the optimal disturbances in Proposition 1, then present the robust filter in Proposition 2, and finally discuss parameter estimation and inference.

4.1. Optimal Disturbances
To derive the disturbances without making distributional assumptions, we apply minimax control theory

(see, e.g., Zames 1981, Başar and Bernhard 1995). Using (3), we first re-express $L(\alpha_t, v_t, \omega_t) < \gamma$ as follows:

$$J(\alpha_t, \omega_t) = \sum_{t=1}^{N} \|\alpha_t - \hat{\alpha}_t\|^2$$
$$- \gamma \left[\sum_{t=0}^{N-1} (\|\omega_t\|_{Q_t^{-1}}^2 + \|y_t - Z_t \alpha_t\|_{H_t^{-1}}^2) \right.$$
$$\left. + \|\alpha_0 - \hat{\alpha}_0\|_{P_0^{-1}}^2 \right], \qquad (5)$$

where we substituted $v_t = y_t - Z_t \alpha_t$ from Equation (1). Equation (5) expresses the loss function for the statistician and Nature. The first term on its RHS reflects the loss a statistician incurs in estimating the unobserved state, while the second term in the square parentheses captures Nature's impedance via $(\{\omega_t\}, \alpha_0)$.

The resulting game is zero-sum. To see this point, let $J(\alpha_t, \omega_t) = J_1 = -J_2$, where $J_1 = J(\alpha_t, \omega_t)$ denotes the statistician's payoff function, and $J_2 = -J(\alpha_t, \omega_t)$ represents Nature's payoff function. Consequently $J_1 + J_2 = J(\alpha_t, \omega_t) + (-J(\alpha_t, \omega_t)) = 0$ for every (α_t, ω_t), not just at the equlibrium solutions, because of the common objective function $J(\alpha_t, \omega_t)$ in Equation (5) for both players. Also, the value of the game

$$J^* = J(\alpha_t^*, \omega_t^*) = \underbrace{\min}_{\alpha_t} \underbrace{\max}_{\omega_t} J(\alpha_t, \omega_t)$$

is not zero at equilibrium.

Next, to dynamically optimize $J(\alpha_t, \omega_t)$, we apply Pontryagin's maximum principle (see, e.g., Sethi and Thompson 2000). To this end, we re-express the state dynamics in Equation (2) as follows:

$$\Delta \alpha_t = \alpha_{t+1} - \alpha_t = d_t - (I - T_t)\alpha_t + \omega_t. \qquad (6)$$

Then, using (5) for a given t, we formulate the Hamiltonian function

$$\mathcal{H}_t = \|\alpha_t - \hat{\alpha}_t\|^2 - \gamma \left[\|\omega_t\|_{Q_t^{-1}}^2 + \|y_t - Z\alpha_t\|_{H_t^{-1}}^2 \right]$$
$$+ \mu_{t+1}' (d_t - (I - T_t)\alpha_t + \omega_t), \qquad (7)$$

where we have adjoined the RHS of Equation (6) using the $m \times 1$ vector of costate variables, μ_{t+1}. The Hamiltonian $\mathcal{H}_t$ decomposes the dynamic optimization problem $\arg \max J(\alpha_t, \omega_t)$ in (4) into instantaneous static optimization problems in (7). Then we connect the optimized solutions over time using the evolution of costate variables prescribed by Pontryagin's maximum principle (see, e.g., Sethi and Thompson 2000).

Equation (7) has the following interpretation. The first term on the RHS represents the loss in forecast accuracy; the second term accounts for the adverse impact due to the shocks; and the third term captures the long-term impact due to fluctuations in the true state. The larger the costate, the larger the impact of state fluctuations on forecast accuracy. In other words,

the costate variables quantify the difficulty in improving the forecast accuracy due to the fluctuating states.

Based on Pontryagin's maximum principle, the first-order conditions (FOCs) are given by

$$\frac{\partial \mathcal{H}_t}{\partial \omega_t'} = 0 \quad \text{and} \quad \Delta \mu_t = -\frac{\partial \mathcal{H}_t}{\partial \alpha_t}, \tag{8}$$

where $\Delta \mu_t = \mu_{t+1} - \mu_t$. Differentiating $\mathcal{H}_t$, we get $\partial \mathcal{H}_t / \partial \omega_t = -\gamma(\partial(\omega_t' Q_t^{-1} \omega_t)/\partial \omega_t) + \mu_{t+1}' = -2\gamma \omega_t' Q_t^{-1} + \mu_{t+1}'$. Setting the last expression to zero, we get $2\gamma \omega_t^{*\prime} Q_t^{-1} = \mu_{t+1}'$, which yields the optimal disturbances

$$\omega_t^* = Q_t \lambda_{t+1}, \tag{9}$$

where we have re-expressed the costate variables as $\lambda_{t+1} = \mu_{t+1}/(2\gamma)$.

Using Equation (8) we next obtain the costate dynamics. Specifically, $\Delta \mu_t' = -\partial \mathcal{H}_t / \partial \alpha_t' = -2(\alpha_t - \hat{\alpha}_t)' + 2\gamma(y_t - Z_t \alpha_t)' H_t^{-1}(-Z_t) + \mu_{t+1}'(I - T_t)$. Dividing by 2γ, replacing $\mu_t/(2\gamma)$ by λ_t, and taking the transpose, we get $\lambda_{t+1} - \lambda_t = -(\alpha_t - \hat{\alpha}_t)/\gamma - Z_t' H_t^{-1}(y_t - Z_t \alpha_t) + \lambda_{t+1} - T_t'\lambda_{t+1}$. Hence the costate vector evolves as follows:

$$\lambda_t = T_t'\lambda_{t+1} + (\alpha_t - \hat{\alpha}_t)/\gamma + Z_t' H_t^{-1}(y_t - Z_t \alpha_t). \tag{10}$$

Collecting (9) and (10), we obtain the following proposition.

Proposition 1. *The vector of optimal disturbances is given by*

$$\omega_t^* = Q_t \lambda_{t+1},$$

where the costate vector evolves according to

$$\lambda_t = T_t'\lambda_{t+1} + (\alpha_t - \hat{\alpha}_t)/\gamma + Z_t' H_t^{-1}(y_t - Z_t \alpha_t),$$

starting backwards from the terminal $\lambda_N = 0$ for $t = \{N - 1, \ldots, 1, 0\}$.

Equations (9) and (10) provide a mechanism to generate the sequence of optimal disturbances without requiring distributional assumptions. The costate vector is driven by the forward dependence $(T_t'\lambda_{t+1})$, the state estimation error $(\alpha_t - \hat{\alpha}_t)$, and the measurement noise $(y_t - Z_t \alpha_t)$. As γ decreases, the magnitude of the costate vector increases and so does the size of the worst case disturbances ω_t^*. Incorporating these disturbances in the state vector, we next derive the optimal state estimator $\hat{\alpha}_t$.

4.2. Robust Filter

At initial $t = 0$, we maximize $J(\alpha_0, \omega_0)$ in Equation (5) subject to the constraint in Equation (2). Specifically, we maximize the Lagrangian $l(\alpha_0) = -\gamma \|\alpha_0 - \hat{\alpha}_t\|_{P_0^{-1}}^2 + \mu_0'\alpha_0 = -\gamma(\alpha_0 - \hat{\alpha}_t)' P_0^{-1}(\alpha_0 - \hat{\alpha}_t) + \mu_0'\alpha_0$ to get $\partial l / \partial \alpha_0' = -2\gamma(\alpha_0 - \hat{\alpha}_t)' P_0^{-1} + \mu_0'$. Setting the last equality to zero, we get $\alpha_0 = \hat{\alpha}_0 + P_0 \lambda_0$, where $\lambda_0 = \mu_0/(2\gamma)$. We next use this relationship between the true and estimated

states for $t \geq 1$, i.e., $\alpha_t = \hat{\alpha}_t + P_t \lambda_t$. Substituting it and the disturbances $\omega_t^* = Q_t \lambda_{t+1}$ from Proposition 1 in the state Equation (2), we get

$$\hat{\alpha}_{t+1} + P_{t+1} \lambda_{t+1} = d_t + T_t \hat{\alpha}_t + T_t P_t \lambda_t + Q_t \lambda_{t+1}. \tag{11}$$

Similarly, we substitute $\alpha_t = \hat{\alpha}_t + P_t \lambda_t$ in the costate Equation (10) to obtain

$$\lambda_t = T'\lambda_{t+1} + P_t \lambda_t/\gamma + Z_t' H_t^{-1}(y_t - Z_t(\hat{\alpha}_t + P_t \lambda_t)). \tag{12}$$

Then, collecting all of the λ_t terms in (12) to its left-hand side and taking the inverse, we get

$$\lambda_t = \left(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t\right)^{-1} \cdot \left(T_t'\lambda_{t+1} + Z_t' H_t^{-1}(y_t - Z_t \hat{\alpha}_t)\right). \tag{13}$$

Next, we substitute λ_t from (13) in (11), collect the terms for λ_{t+1}, and apply Lagrange's method of undetermined coefficients to observe that

$$\hat{\alpha}_{t+1} = d_t + T_t \hat{\alpha}_t + T_t P_t \left(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t\right)^{-1} \cdot Z_t' H_t^{-1}(y_t - Z_t \hat{\alpha}_t), \tag{14}$$

and

$$P_{t+1} = T_t P_t \left(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t\right)^{-1} T_t' + Q_t. \tag{15}$$

Collecting (14) and (15), we state the following proposition.

Proposition 2. *The optimal state estimator $\hat{\alpha}_t$ is given recursively for $t = 1, \ldots, N$ by*

$$\hat{\alpha}_{t+1} = d_t + T_t \hat{\alpha}_t + T_t K_t(y_t - Z_t \hat{\alpha}_t), \tag{16}$$

where

$$K_t = P_t(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t)^{-1} Z_t' H_t^{-1}, \quad \text{and} \tag{17}$$

$$P_{t+1} = T_t P_t(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t)^{-1} T_t' + Q_t. \tag{18}$$

Equations (16) through (18) provide the optimal recursions to update the state estimate $\hat{\alpha}_{t+1}$ based on $\hat{\alpha}_t$. This update is proportional to the estimation error $(y_t - Z_t \hat{\alpha}_t)$. The proportionality depends on the gain matrix K_t. The gain matrix should satisfy the condition that the matrix $(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t)$ is positive definite to ensure optimality (i.e., the second-order conditions).

The subtraction of P_t/γ in (17) *optimally* expands the gain matrix K_t. To see this clearly in (17), observe that as γ decreases, the matrix P_t/γ increases, so $(-P_t/\gamma)$ decreases, which in turn increases $(I - P_t/\gamma + Z_t' H_t^{-1} Z_t P_t)^{-1}$ due to the inverse operation, and hence K_t increases. Consequently, a larger K_t places greater weight on the current market data $(y_t - Z_t \hat{\alpha}_t)$. Thus, a more robust filter overweighs the recent past, making it relatively more "present-oriented." Similar reasoning shows that $P_{t+1}(\gamma)$ in (18) increases as γ

decreases. As a result, a more robust filter also accommodates a wider range of perturbations to the unobserved states α_t.

In Section B of the web appendix, we prove the robust filter converges to the Kalman filter as $\gamma \to \infty$. Specifically, we show that the gain matrices, the covariance matrices, and the state estimators of the standard Kalman filter and the proposed robust filter are asymptotically equivalent. Furthermore, the robust filter becomes less conservative as γ increases and so it is the least conservative filter at $\gamma \to \infty$. Thus, the Kalman filter is not only nested but also the least conservative filter.

4.3. Robust Parameter Estimation and Inference

The time-varying system matrices $\{Z_t, T_t, d_t, H_t, Q_t\}$ in Equations (1) and (2) depend on the $p \times 1$ vector of model parameters θ. To estimate θ, as in the Kalman filter estimation, we maximize the criterion

$$S_\gamma(\theta) = -\tfrac{1}{2} \sum_{t=1}^{N} \left[\ln(\det(F_t(\theta))) \right.$$
$$\left. + (y_t - Z_t \hat{\alpha}_t(\theta))' F_t(\theta)^{-1}(y_t - Z_t \hat{\alpha}_t(\theta)) \right], \quad (19)$$

where $F_t(\theta) = Z_t P_t(\theta) Z_t' + H_t(\theta)$. More precisely, for a given level of conservatism γ, we obtain the robust parameter estimates

$$\hat{\theta} = \arg\max S_\gamma(\theta). \quad (20)$$

To conduct robust inference, we extract the square root of the diagonal of the negative inverse of the Hessian of $S_\gamma(\theta)$ to obtain the vector of robust standard errors

$$se(\hat{\theta}) = \mathrm{Sqrt}(\mathrm{Diag}(-\hat{G}^{-1})), \quad (21)$$

where the $p \times p$ Hessian matrix $\hat{G} = (\partial^2 S_\gamma(\theta)/\partial\theta\partial\theta')|_{\theta=\hat{\theta}}$ is evaluated at the estimated parameter values.

In summary, we derived the optimal disturbances in Equation (9) without making any distributional assumptions; Equations (16) through (18) provide the robust filter recursions; Equation (20) yields the robust parameter estimates; and Equation (21) enables robust inference. These equations, together, complete the proposed approach for robust dynamic estimation shown in Panel C of Figure 1.

To complete the approach mentioned in Panel D of Figure 1, we "correct" the standard errors obtained from Equation (21) via White (1980). Specifically, for each observation t, we first numerically compute the gradient $\partial S_\gamma(\theta)/\partial\theta'$, which is a $1 \times p$ vector. Next, we stack all of the gradients row-wise to create the $N \times p$ matrix A and thus create the matrix $M = A'A$ of dimension $p \times p$. Then, we obtain the vector of double-robust standard errors

$$se(\hat{\theta}) = \mathrm{Sqrt}(\mathrm{Diag}(G^{-1} M G^{-1})). \quad (22)$$

Note that the robust parameter estimates $\hat{\theta}$ (from Equation (20)) remain the same in both standard error estimators (i.e., Equations (21) or (22)). Hence double-robustification affects only the inference, and does not alter the managerial outcomes based on Equation (20).

4.4. When To Use Which Method?

Should researchers or analysts always pursue robust estimation and inferences? Not necessarily. The answer depends on whether the residuals exhibit nonnormality and whether we suspect misspecification. To this end, we suggest using the Jarque–Bera test to ascertain normality and White's test to discover heteroskedasticity or misspecification.

The Jarque–Bera test statistic is given by

$$T_1 = (N - p)(S^2 + 0.25(C - 3)^2)/6, \quad (23)$$

where the sample skewness

$$S = \left(\frac{\sum_{t=1}^{N} (r_t - \bar{r})^3}{\sum_{t=1}^{N} (r_t - \bar{r})^2} \right)^{3/2},$$

the sample kurtosis

$$C = \left(\frac{\sum_{t=1}^{N} (r_t - \bar{r})^4}{\sum_{t=1}^{N} (r_t - \bar{r})^2} \right)^2,$$

and $(r_t, \bar{r})$ denotes the individual residuals and their average. The null hypothesis is that the residuals are normally distributed (i.e., neither skewness nor kurtosis). The statistic T_1 follows χ^2 with two degrees of freedom, which yields the p-values at the 95% confidence level.

To conduct the White's (1980) test, researchers or analysts should run an auxiliary regression with the dependent variable as the squared residuals, and the regressors being the constant plus the linear, quadratic, and cross-product terms of the variables in the original model. Then the White (1980) test statistic is given by

$$T_2 = NR^2, \quad (24)$$

where R^2 denotes the unadjusted fit of the auxiliary regression. The null hypothesis is that neither heteroskedasticity nor misspecification exists. The statistic T_2 follows χ^2_k with k degrees of freedom, which equals the number of regressors excluding the constant, and it yields the p-values at the 95% confidence level.

If T_1 holds, i.e., one fails to reject the null, then the residuals are deemed normal; if T_2 holds, then neither heteroskedasticity nor misspecification exists. Accordingly, researchers or analysts run the Kalman filter and then apply T_1 and T_2 to the residuals from Panel A. If T_1 and T_2 hold, they stay in Panel A. There is no need for robust estimation or robust inferences. If T_1 holds but

not T_2, they go to Panel B and apply the sandwich estimator as in Equation (22), but using the gradient and Hessian matrices obtained from the Kalman filter (for details, see Naik et al. 2008). If T_1 does not hold, they go to Panel C and run the robust filter. Then they apply T_2 to the residuals obtained from the robust filter. If T_2 holds, they stay in Panel C. If not, they go to Panel D and the sandwich estimator given in Equation (22) is applied. In general, this road map guides researchers and analysts to discover for themselves whether they should implement the results from robust estimation and inferences in their particular empirical contexts. Next, we apply the robust dynamic estimation to blood advertising by Canadian Blood Services.

5. Empirical Application
5.1. Blood Marketing
Over a hundred million blood donations are collected worldwide every year. Blood donors give blood without compensation and without knowing who the donees are. Thus, blood donations epitomize altruism, which blood collection agencies elicit through advertising. Yet advertising for blood differs from brand advertising. Unlike conventional marketing, profit considerations do not arise for blood collection agencies since there are no market prices or profit margins involved. Moreover, collection agencies do not advertise to maximize donations but to bring blood collections close to targets since blood cannot be stored for long periods.

The literature on blood donations is sparse. A recent study by Aravindakshan et al. (2014) investigates how online media generates blood donations for a small community blood bank. By contrast, we study advertising by CBS, a large national blood collection agency serving English-speaking Canada.

For historical perspective, recall that the government of Premier Jean Chrétien established the CBS because, during the 1980s, over 30,000 people received tainted blood products due to the negligence of the Canadian Red Cross in performing the necessary tests to detect HIV and Hepatitis C. In response, the Royal Commission, headed by Justice Horace Krever, investigated this nationwide public calamity. Based on the commission's report (Krever 1998), the Canadian government banned the Red Cross from collecting blood. Red Cross blood collection was replaced by the CBS (see www.blood.ca) for collecting and supplying blood in Canada, except Québec. In this new system, Canadian Provinces and Territories provide funding to CBS, which delivers blood products to hospitals free of charge without consideration of the market prices or profit margins. As a state monopoly, CBS does not face competition for blood collection in any of its five provinces, i.e., Alberta, the Atlantic Provinces, British Columbia, Ontario, and the other Prairies. Moreover,

based on the Krever Commission, CBS cannot import blood from outside these five provinces, and the hospitals cannot receive blood products from other sources.

5.2. Data Description
CBS deploys marketing communications to elicit altruistic behaviors from donors. Consistent with its motto, *It's in you to give*, CBS advertises to motivate people to give blood by using mass communications via radio advertisements and personal communications via phone calls. We collected weekly data on pints of blood collected, radio gross rating points (GRPs), and the number of phone calls for each of the five provinces from April 2010 through March 2013. Thus, we observed the entire blood collection in English-speaking Canada for 156 weeks. Table T1 in the web appendix (see Section A) summarizes the descriptive statistics.

5.3. Blood Collection Dynamics
Let y_t denote the actual blood collected at time t, b_t denote the mean blood levels, and v_t represent the disturbance such that

$$y_t = b_t + \sqrt{h}v_t, \tag{25}$$

where h measures the impact of the disturbances v_t. The mean blood level evolves over time as follows:

$$b_{t+1} = \beta_1\sqrt{u_{1t}} + \beta_2\sqrt{u_{2t}} + \phi b_t + \sqrt{q}\omega_t, \tag{26}$$

where (u_{1t}, u_{2t}) denotes the weekly radio GRPs and the number of phone calls, (β_1, β_2) correspond to their effectiveness, ϕ represents the carryover effect, and q measures the impact of the disturbance ω_t. In robust estimation, the disturbances (v_t, ω_t) are conceptually viewed as Nature's perturbations that are strategically set rather than randomly drawn (as in the nonrobust Kalman filter estimation). Consequently, $(\sqrt{h}, \sqrt{q})$ are interpreted as the impact of marginal changes in (v_t, ω_t) on the observation and state variables (y, b), respectively. Formally, $\sqrt{h} = \partial y/\partial v$ and $\sqrt{q} = \partial b/\partial \omega$. Thus, they represent the impacts of disturbances (v, ω) on (y, b), respectively.

Equations (25) and (26) can be cast into the dynamic linear model in (1) and (2), respectively. To see this equivalence, note that $Z_t = 1$, $\alpha_t = b_t$, $H_t = h$, $T_t = \phi$, $d_t = \beta_1\sqrt{u_{1t}} + \beta_2\sqrt{u_{2t}}$, and $Q_t = q$. We observe (y_t, u_{1t}, u_{2t}) for $t = 1, \ldots, N = 156$ weeks. We must estimate the parameter vector $\theta = (\beta_1, \beta_2, \phi, h, q)'$ of dimension $p = 5$. We use 104 observations for estimation and hold out 52 observations for validation. We next apply the four methods mentioned in Panels A–D of Figure 1. For details of Panels A and B, see Naik et al. (1998, 2008), respectively. For details of Panels C and D, see Sections 4.2 and 4.3.

5.4. Scale for Conservatism

Here we create a scale such that conservatism increases as κ increases rather than as γ decreases (i.e., reverse the orientation). We first estimate the Kalman filter to obtain the parameter estimates and the maximized value $S^*_{\gamma=\infty}$ in Equation (19). Then we apply the robust filter to estimate the model for various values of γ, and select that value of γ, which we denote by $\hat{\gamma}_{\min}$, for which the difference $-2(S^*_\gamma - S^*_{\gamma=\infty})$ just equals $\chi^2_{1,0.05} = 3.84$. In words, we conceptually seek that robust filter that is statistically no worse than the Kalman filter. For Alberta, Atlantic, British Columbia, Ontario, and the Prairies, we find $\hat{\gamma}_{\min} = (4{,}000, 3{,}000, 3{,}000, 50{,}000, 4{,}000)$, respectively. Next, we increase γ in increments of 1,000 up to a large value (which occurs at 103,000 in our application) such that the parameter estimates from the robust filter converge to those obtained from the Kalman filter. Finally, we create a scale $\kappa = \exp(-(\gamma - \hat{\gamma}_{\min})/\hat{\gamma}_{\min})$ such that conservatism increases as κ increases (i.e., reverse the orientation with respect to γ). In this scale, κ and γ map one-to-one with $\kappa = 0$ corresponding to the least conservative filter and $\kappa = 1$ to the most conservative filter.

5.5. Parameter Estimates

We estimated 500 models for various γ and 5 provinces. Table 1 reports the parameter estimates and t-values across the four approaches in Figure 1. Panels A and B of Table 1 present the parameter estimates and their t-values from the Kalman filter without and with White's correction, respectively. The estimated parameter values in Panels A and B do not change; only the t-values differ because the sandwich estimator alters the standard errors to hedge misspecification risk.

By contrast, Panel C of Table 1 displays the results from the proposed robust filter. Recall that the robust filter uses the optimal disturbances derived in Proposition 1, eschewing probabilistic assumptions for the random errors. Consequently, the estimated parameter values obtained via the Equation (20) for the robust filter (Panel C) differ from those obtained via the Kalman filter (Panel A). This outcome is the consequence of parametric robustification, which places a greater weight on current market data compared to the Kalman filter (i.e., present-orientedness of the robust filter). In addition, the standard errors from Equation (21) for the robust filter (Panel C) differ from those obtained via the Kalman filter (Panel A). Thus, we obtain the robust estimates *and* robust inferences.

To achieve "double-robust" inferences, we use the sandwich estimator in Equation (22). The parameter values in Panels C and D do not change since both are obtained via Equation (20). Only the t-values differ because Equation (21) yields the standard errors for the robust filter, whereas Equation (22) yields the standard errors from the sandwich estimator. However, double robustification in Panel D alters the parameter estimates (relative to Panel B) and their significance (relative to Panels A–C).

Substantively, the results show that the carryover effects are significant in all provinces except Atlantic and they appear bi-modal with values clustered in the neighborhoods of 0.5 and 0.9. We also learn that radio advertising is statistically significant in all provinces, except Atlantic where phone calls generate blood donations. Phone call communications are statistically significant in all provinces, except British Columbia where radio advertising generates blood donations. Moreover, heterogeneity exists across provinces. The effectiveness of radio advertising estimated with the robust filter ranges from 0.06 to 1.23 with a median of 0.83. Similarly, the phone call effectiveness ranges from 0.23 to 1.47 with a median of 0.88.

5.6. Compensatory Effect

Figure 2 presents the effect of conservatism on the carryover effect, radio effectiveness, and phone call effectiveness for Alberta. As conservatism increases, we

Table 1. Parameter Estimates (t-Values)

Provinces	Radio $\hat{\beta}_1$	Calls $\hat{\beta}_2$	Carryover $\hat{\phi}$	Radio $\hat{\beta}_1$	Calls $\hat{\beta}_2$	Carryover $\hat{\phi}$
	Panel A. Kalman filter			Panel B. White's correction		
Alberta	0.90 (2.15)	0.44 (2.17)	0.88 (18.23)	0.90 (2.30)	0.44 (2.21)	0.88 (19.20)
Atlantic	0.06 (0.29)	1.33 (6.08)	0.28 (1.43)	0.06 (0.29)	1.33 (4.72)	0.28 (1.05)
B.C.	0.66 (2.72)	0.20 (1.60)	0.93 (29.43)	0.66 (2.91)	0.20 (1.62)	0.93 (27.07)
Ontario	1.05 (1.73)	0.71 (2.65)	0.87 (17.76)	1.05 (1.85)	0.71 (2.35)	0.87 (15.13)
Prairies	0.78 (2.31)	0.98 (4.79)	0.62 (6.49)	0.78 (2.50)	0.98 (3.95)	0.62 (5.24)
	Panel C. Robust filter			Panel D. Combined approach		
Alberta	1.07 (2.44)	0.61 (2.91)	0.83 (15.63)	1.07 (2.47)	0.61 (2.77)	0.83 (15.45)
Atlantic	0.06 (0.32)	1.47 (8.58)	0.15 (0.82)	0.06 (0.31)	1.47 (6.79)	0.15 (0.60)
B.C.	0.72 (2.87)	0.23 (1.73)	0.91 (26.88)	0.72 (2.96)	0.23 (1.63)	0.91 (23.68)
Ontario	1.23 (1.97)	0.88 (3.19)	0.83 (15.79)	1.23 (2.08)	0.88 (2.54)	0.83 (12.31)
Prairies	0.82 (2.42)	1.21 (6.63)	0.51 (13.87)	0.82 (2.57)	1.21 (5.50)	0.51 (4.27)

Figure 2. (Color online) Compensatory Effect of Conservatism
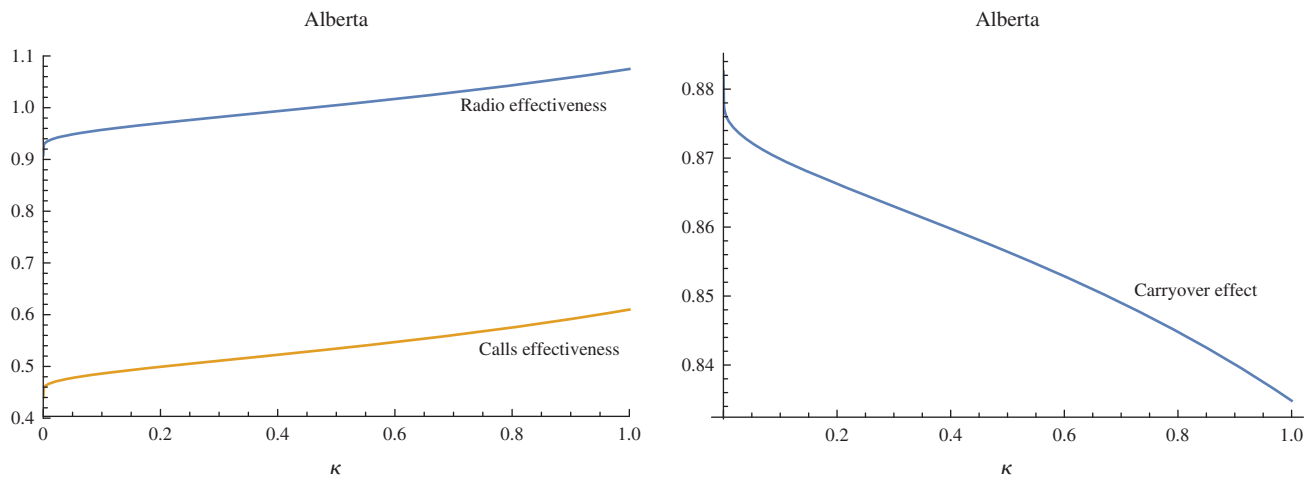
observe that the estimated effectiveness of radio advertising increases; the estimated effectiveness of phone calls also increases; whereas the estimated carryover effect decreases. These findings generalize across all five provinces.

These results reveal the compensatory effect between marketing effectiveness and the carryover effect as conservatism increases. That is, marketing effectiveness and the carryover effect move in the opposite direction. Thus, conservatism induces managers to rely more on the present and less on the past. We attribute this finding to the present-orientedness of the robust filter whose gain matrix places greater weight on the current data, thereby increasing short-term effectiveness and decreasing long-term carryover effect.

5.7. Marketing Elasticities

To compare the relative performance of CBS' marketing instruments, we compute the elasticities for radio advertising and personal phone calls. Specifically, for each province, we draw 100,000 realizations from the distribution of estimated parameter values, which takes into account the associated parametric uncertainty and their cross-correlations via the full variance-covariance matrix. We obtain the median elasticities and their statistical significance based on the 2.5th and 97.5th percentiles.

Radio advertising elasticities for Alberta, Atlantic, British Columbia, Ontario, and the Prairies are (0.09, 0.002, 0.16, 0.07, 0.03), respectively. All of the elasticities, except Atlantic, are statistically significant. Substantial heterogeneity exists across provinces, ranging from negligible in Atlantic to 0.16 in British Columbia. To provide a perspective, consider the meta-analysis by Sethuraman et al. (2011) based on the extant advertising literature. They present elasticities for television, print, and aggregate advertising, but *no elasticity for*

radio advertising is reported in their meta-analysis, indicating its scant availability in the extant literature. Furthermore, studies in the meta-analysis focus on industries such as pharmaceutical, durable goods, food, nonfood, and services, where consumers receive tangible benefits from the advertised brand. By contrast, blood donors do not receive tangible benefits. Despite these differences, it is remarkable that the median elasticity of radio advertising for blood is 0.07, which is comparable to the median advertising elasticity of 0.10 (ibid, p. 468). Thus, our work augments this scant literature by contributing the elasticities of radio advertising.

To complement mass advertising on radio, CBS also uses its National Contact Center to make phone calls, which constitute one-on-one personal communications. The elasticities of personal communications for Alberta, Atlantic, British Columbia, Ontario, and the Prairies are (0.33, 0.27, 0.27, 0.35, 0.29), respectively. All of the elasticities, except for British Columbia, are statistically significant. Unlike radio advertising, much heterogeneity does *not* exist across provinces, with the elasticities ranging from 0.27 to 0.35. To provide a perspective, consider the meta-analysis by Albers et al. (2010) based on the extant literature on personal selling. Over 50% of their studies cover the pharmaceutical industry (e.g., sales force communication to physicians) and about 25% pertain to military recruiting efforts. *No elasticity for personal communications of nonprofit organizations is reported in their meta-analysis, again indicating scant availability in the extant literature.* Despite this difference, it is remarkable that the mean elasticity for phone call communications for blood donations is 0.29, which is comparable to the mean elasticity for personal selling of 0.31 (ibid, p. 840). Thus, we further augment the scant literature by contributing the elasticities of phone call communications.

5.8. Was Robustification Necessary?

We answer this question based on the road map given in Section 4.4. Specifically, we trace the diagnostic path in our empirical context, starting with Kalman filter residuals in Panel A of Figure 1, applying the Jarque–Bera test, and obtaining $T_1 = 0.707$ (p-value = 0.001), and thus rejecting the normality assumption. Because T_1 does not hold, we go to Panel C of Figure 1, conduct robust dynamic estimation, apply White's test to the residuals from the robust filter, obtain the test statistic $T_2 = 123.88$ (p-value = 0.000), and thus reject homoscedasticity and no misspecification. Because T_2 does not hold, we go to Panel D of Figure 1 and apply the sandwich estimator to achieve double robustification of inferences. Hence, our market data support use of the combined approach in Panel D for robust parameter estimation and doubly robust inferences.

5.9. Fit vs. Forecasts Trade-off

Robustification induces a trade-off between in-sample fit and out-of-sample forecasts. To understand this, we use 104 observations to fit the model and hold out 52 observations to assess predictive accuracy. Table 2 reports the results. Panel A shows that the Kalman filter better fits the sample data (i.e., larger log-likelihood values) for all five provinces. By contrast, Panel B shows that the robust filter uniformly outperforms the Kalman filter on the out-of-sample predictions across the three metrics, i.e., mean squared error (MSE), mean absolute percentage error (MAPE), and mean absolute deviation (MAD), in all five provinces. This finding suggests that the out-of-sample data degrade the Kalman filter's performance much more than the robust filter's performance. The intuition is as follows.

Table 2. Fit vs. Forecasts Trade-off

Panel A: In-sample fit		
	Log-likelihood	
Provinces	Kalman filter	Robust filter
Alberta	−464.77	−466.68
Atlantic	−450.07	−450.66
B.C.	−418.30	−419.77
Ontario	−580.75	−582.67
Prairies	−471.32	−472.89

Panel B: Out-of-sample prediction					
	MSE		MAPE (%)		MAD
Provinces	Kalman filter	Robust filter	Kalman filter	Robust filter	Kalman filter	Robust filter
Alberta	846.1	532.5	3.9	3.1	22.1	17.4
Atlantic	1,752.1	1,632.9	8.6	8.3	32.2	31.2
B.C.	501.2	394.9	3.7	3.3	18.3	16.4
Ontario	9,636.7	6,791.1	4.5	3.8	70.3	59.5
Prairies	1,193.3	893.8	6.7	5.8	27.5	24.0

In the estimation sample, the Kalman filter performs better than the robust filter because, by construction, the latter sacrifices up to 3.84 points on the χ^2-scale to eschew distributional assumptions; its benefit accrues in the holdout sample when the data generation process departs from normality or the assumed model is misspecified.

6. Dynamic Multimedia Allocation Under Conservatism

Given the robust estimation results, we tackle the budget allocation problem of the large public health agency whose objective differs from profit-maximization. Specifically, we seek insights into how managers should determine the optimal marketing budget and optimally allocate it to radio advertising and phone call communications. We investigate these issues next.

CBS manages each province independently and sets the collection targets. A deviation from the targets is costly because blood levels below the target create shortages, whereas levels above it generate waste (e.g., Brodheim and Prastacos 1979, Prastacos 1984). Because the cost of excess or shortage is asymmetric, we use the LINEX function (Varian 1975, Zellner 1986),

$$g(b, \tau) = e^{-a(b-\tau)} + a(b - \tau) - 1, \tag{27}$$

where (b, τ) are the actual and target blood levels and $a > 0$ controls the magnitude of asymmetry. In Equation (27), $g(\cdot)$ represents the cost that exponentially increases as the shortage increases (i.e., $b - \tau$ become more negative) due to the first term on the RHS; the cost grows linearly when the excess blood level increases (i.e., $b - \tau$ become more positive) due to the second term on the RHS; the third term ensures that the cost is zero when $b = \tau$.

CBS allocates its marketing budget $B(t)$ to radio advertising and phone calls to drive blood collections. To minimize the total cost of excess and shortage plus the marketing budget $B(t)$, we specify the objective function

$$f(u_1, u_2) = \mathbb{E}\left[\int_0^\infty e^{-\rho t}\left\{e^{-a(b-\tau)} + a(b - \tau) - 1 + B(t)\right\} dt\right], \tag{28}$$

where ρ is the discount rate, and the blood collection evolves according to

$$db = \left(\beta_1\sqrt{u_1} + \beta_2\sqrt{u_2} - \delta b\right) dt + \sqrt{q}\, dW. \tag{29}$$

Equation (29) is an equivalent continuous-time version of the discrete-time blood dynamics in Equation (26). Comparing the two equations, we note that $dW = \omega_t\, dt$ denotes the increment of a standard Brownian motion, $\delta = 1 - \phi$, and $u_1(t)$ and $u_2(t)$ are the radio GRPs and the

number of phone calls. The marketing budget equals $B(t) = c_1 u_1(t) + c_2 u_2(t)$, where c_1 and c_2 are the costs per GRP and phone call, respectively.

We want to identify the optimal strategies (u_1^*, u_2^*) that minimize the total expected cost in Equation (28). To this end, we define the value function

$$V(b) = \min_{(u_1, u_2)} \{f(u_1, u_2)\},$$

which satisfies the Hamilton–Jacobi–Bellman (HJB) equation

$$\rho V(b) = \min_{(u_1, u_2)} \left\{ [e^{-a(b-\tau)} + a(b-\tau) - 1 + B(t)] + V_b(\beta_1 \sqrt{u_1} + \beta_2 \sqrt{u_2} - \delta b) + q V_{bb}/2 \right\}, \quad (30)$$

where $V_b = \partial V/\partial b$ and $V_{bb} = \partial^2 V/\partial b^2$.

Equation (30) has three terms on its RHS. The first term $[e^{-a(b-\tau)} + a(b-\tau) - 1 + B(t)]$ represents the direct cost incurred every period. The second term, $V_b(\beta_1 \sqrt{u_1} + \beta_2 \sqrt{u_2} - \delta b)$, consists of two parts, i.e., (i) the expected incremental blood collected $\mathbb{E}[db/dt] = \beta_1 \sqrt{u_1} + \beta_2 \sqrt{u_2} - \delta b$ due to the marketing activities (u_1, u_2) and (ii) the valuation V_b of a pint of incremental blood. Thus, $V_b(\beta_1 \sqrt{u_1} + \beta_2 \sqrt{u_2} - \delta b)$ adds the indirect contribution to the value function. The third term, $q V_{bb}/2$, accounts for the dynamic uncertainty in blood collection, $\text{Var}[db/dt]$.

Equation (30) is a second-order partial differential equation (PDE). To solve this PDE, we first differentiate its RHS with respect to (u_1, u_2) to obtain the FOCs and then substitute them back into (30) to eliminate the minimization operator. Because blood collection operates close to the target, we also expand $g(b, \tau)$ by the Taylor series around τ and use Lagrange's method of undetermined coefficients to fully characterize the value function as follows:

$$V(b) = A_2(b-\tau)^2/2 + A_1(b-\tau) + A_0, \quad (31)$$

where the coefficients (A_0, A_1, A_2) depend on the model parameters:

$$A_2 = \frac{2\delta + \rho - \sqrt{(\rho + 2\delta)^2 + 2a^2(\beta_1^2/c_1 + \beta_2^2/c_2)}}{\beta_1^2/c_1 + \beta_2^2/c_2},$$

$$A_1 = \frac{2\tau\delta(-(\rho + 2\delta) + \sqrt{(\rho + 2\delta)^2 + 2a^2(\beta_1^2/c_1 + \beta_2^2/c_2)})}{(\beta_1^2/c_1 + \beta_2^2/c_2)(\rho + \sqrt{(\rho + 2\delta)^2 + 2a^2(\beta_1^2/c_1 + \beta_2^2/c_2)})}$$

and

$$A_0 = \frac{3(\beta_1^2/c_1 + \beta_2^2/c_2)(A_1 - \tau A_2)^2}{4\rho} + \frac{q A_2 + \tau(2\rho A_1 + \tau(a^2 - A_2\rho))}{2\rho}.$$

Note that A_1 is always positive and that A_2 is always negative. Having analytically solved the PDE in (30), we present the optimal marketing strategies in Proposition 3.

6.1. Optimal Marketing Resource Allocations
Proposition 3. *The optimal marketing strategies are given by*

$$u_i^* = \begin{cases} (K_{0i} - K_{1i}(b-\tau))^2 & \text{if } b < \tilde{b}, \\ 0 & \text{otherwise}, \end{cases} \quad (32)$$

for $i = (1, 2)$*, where*

$$K_{0i} = \frac{A_1 \beta_i}{2c_i} > 0, \qquad K_{1i} = -\frac{A_2 \beta_i}{2c_i} > 0,$$

and $\tilde{b}$ *denotes the buffer level corresponding to the zero optimal marketing spends.*

Equation (32) has the following properties. First, they are feedback (or closed-loop) strategies as a function of the state variable b. Second, they exhibit nonlinear state dependence due to $(b-\tau)^2$. Last, they operate so as to restore the target level τ.

To understand this last point, consider what happens when blood collection exceeds the target, i.e., $b > \tau$. Because $(b-\tau) > 0$, optimal advertising should be reduced per (32) as b increases, which in turn suppresses the rate of blood collection due to Equation (29). On the other hand, when blood collection falls below the target, i.e., $b < \tau$, optimal advertising should be increased because $(b-\tau) < 0$ in Equation (32), which in turn boosts the blood collection rate due to (29) and thus restores collection levels closer to the target. This mechanism to advertise more (less) when the state variable is small (large) is the *opposite* of the advertising-to-sales ratio heuristic, which recommends spending more (less) when sales are high (low), as used by for-profit companies (see, e.g., Farris et al. 1998).

6.2. Buffer Levels
Should forward-looking managers stop advertising when the collection reaches the target or continue advertising beyond the target? Proposition 3 shows that optimal advertising stops when $u_i^* = 0$, which corresponds to the buffer level $\tilde{b} = \tau(1 + \eta)$. This buffer level *always* exceeds the target because

$$\eta = \delta \frac{-\rho + \sqrt{(2\delta + \rho)^2 + 4(\beta_1^2/c_1 + \beta_2^2/c_2)}}{2(\beta_1^2/c_1 + \beta_2^2/c_2 + \delta(r + \delta))} > 0.$$

Hence, forward-looking managers should *not* stop advertising *even if they reach the target collections*. Why? Because if they were to stop advertising at the target, the *expected* collection levels may soon fall below the target due to Equation (29), thus creating shortages. Anticipating this impending shortfall and responding to prevent it, the manager must continue to advertise to build a healthy buffer.

To provide further guidance on how much buffer level is "healthy," we empirically evaluate the optimal

buffer levels by taking 10,000 draws from the distribution of estimated parameters along with asymmetry parameter $a = 1$ and the costs $c_1 = \$16$ and $c_2 = \$2.2$ obtained from the CBS' marketing team and $\rho = 5\%$ per annum discount rate. Results show that the optimal buffer level should be 44% larger than the actual target in Alberta; 77% larger in Atlantic; 43% larger in British Columbia; 33% larger in Ontario; and 65% larger in the Prairies. On average, CBS should maintain a 36% buffer over their collection targets.

6.3. Budget and Allocation
How does the pursuit of conservatism affect the marketing budget? To answer this question, we evaluate Equation (32) for the optimal radio GRPs and phone calls in each of the five provinces. Table 3 reports the average weekly optimal marketing budget in each province. The optimal conservative and nonconservative budgets are about the same in the Prairies. However, the optimal conservative budget exceeds the nonconservative budget in Alberta by 7.8%, in British Columbia by 2.5%, and in Ontario by 4.6%. By contrast, the optimal conservative budget is below the nonconservative budget in Atlantic by 5%. Across the five provinces, the total conservative budget exceeds the nonconservative budget by 3.7%. Yet the optimal conservative budget is 54.4% below the actual budget; this shows that CBS overspends on marketing.

Besides overspending, how do they allocate the budget to radio versus phone advertising? For each province, Table 4 presents the optimal budget allocation to radio advertising as a ratio of the total optimal budget, i.e., $c_1 u_1^* / B^*$. The optimal allocation to

radio advertising varies substantially across provinces from 0% to 56% with a median of 21%. Because the remaining budget is allocated to phone calls, the phone calls get (72%, 100%, 44%, 79%, 95%) in Alberta, Atlantic, British Columbia, Ontario, and the Prairies, respectively.

The median phone call allocation of 79% compared to radio advertising at 21% comports with their relative elasticities. Specifically, our estimation results (see Section 5.7) show that phone calls (radio advertising) elasticity is 0.29 (0.07). That is, *personal phone calls are four times more effective than radio advertising.* However, the heuristic of allocating budgets proportional to the relative elasticities does not uniformly apply to all provinces. For example, this heuristic would allocate 37% to radio advertising in British Columbia, whereas the optimal allocation recommends 56% (via Equation (32)), thereby resulting in a sub-optimal under-allocation to radio advertising.

6.4. Profit vs. Volatility Trade-off
Robustification induces another trade-off: The total profit reduces, but so does its volatility. Conservatism influences the long-term profit level and its variability *differentially*. For this insight, we compute the expectation of the value function for various levels of conservatism. We represent the value function as the negative of the total cost function and interpret it as the long-term profit. That is, we evaluate and plot $(\mathbb{E}(V)_{\kappa=0} - \mathbb{E}(V)_{\kappa=\kappa})/\mathbb{E}(V)_{\kappa=0}$ for different values of κ from zero to unity. This scaled expectation measures the percentage *profit reduction* as a function of conservatism. Similarly, we compute the variance of the value function $V(b)$, i.e., $(\text{var}(V)_{\kappa=0} - \text{var}(V)_{\kappa=\kappa})/\text{var}(V)_{\kappa=0}$. This scaled variance quantifies the percentage *volatility reduction* as a

Table 3. Optimal Weekly Budgets

	Kalman filter ($\kappa = 0$) (\$)	Robust filter ($\kappa = 1$) (\$)
Alberta	10,662	11,493
Atlantic	7,493	7,121
B.C.	8,152	8,355
Ontario	72,327	75,651
Prairies	8,372	8,346
Total	107,006	110,966

Table 4. Optimal Radio Allocation[a]

	Kalman filter ($\kappa = 0$) (%)	Robust filter ($\kappa = 1$) (%)
Alberta	37	28
Atlantic	0	0
B.C.	60	56
Ontario	22	21
Prairies	8	5
Total	24	22

[a]The remaining budget is allocated to phone call communications.

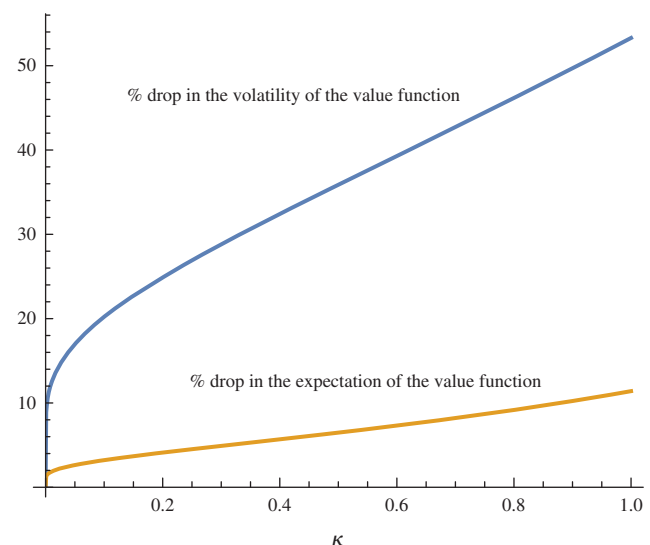**Figure 3.** (Color online) Impact of Conservatism on Profit and Its Variance

Table 5. Optimal Robustness

κ^*	$\rho = 0.003$	$\rho = 0.03$	$\rho = 0.3$
Alberta	0.520	1.000	1.000
Atlantic	0.000	0.000	0.521
B.C.	0.999	1.000	1.000
Ontario	0.999	1.000	1.000
Prairies	0.001	0.169	1.000

function of conservatism. Figure 3 presents the results for the Alberta province. It reveals that, as conservatism increases, the profit reduction increases above 10% at $\kappa = 1$, but the volatility reduction increases more sharply, beyond 50% at $\kappa = 1$. The plots for other provinces (not shown for brevity) are qualitatively similar. This profit-volatility trade-off is reminiscent of the returns-risk trade-off in finance; for example, a more conservative investor can add more bonds to the portfolio and thus reduce returns as well as volatility. To determine optimal conservatism,[1] we combine profit and volatility to formulate $U(\kappa) = \mathbb{E}(V)_\kappa - \rho \operatorname{var}(V)_\kappa$, where ρ denotes the risk aversion coefficient, and thus obtain $\kappa^* = \arg\max_{0 \le \kappa \le 1} U(\kappa)$. We apply NMaximize in *Mathematica* 10.1 and report κ^* for various values of ρ in Table 5. For example, when risk aversion is small $\rho = 0.003$, managers should be least robust in Atlantic and the Prairies, moderately robust in Alberta, and most robust in British Columbia and Ontario.

7. Conclusions

Optimal budgeting and allocation of marketing resources over time require estimation of dynamic models (see, e.g., Gatignon 1993, Mantrala 2002, Shankar 2008, Gupta and Steenburgh 2008). Virtually all estimation methods assume that error terms follow the Normal distribution or some other probability distribution, parametric or nonparametric. When the data generation process differs from the assumed probability distribution, model misspecification arises. The extant literature recognizes the seriousness of this misspecification problem and recommends using the sandwich estimator (White 1980) to "correct" the statistical inferences. However, although the sandwich estimator affects statistical outcomes such as inference, it does *not* alter the managerial actions of budgeting and allocation because the estimated parameters remain unchanged, with or without this correction. Consequently, despite White's (1980) sandwich estimator, managers continue to risk misallocating marketing resources in the presence of misspecification. This was our motivation to develop a method for robust dynamic estimation. We apply minimax control theory to derive the robust filter for a general time-varying multivariate dynamic linear model without making any distributional assumptions. Specifically, we first

derive the optimal disturbances in Proposition 1 via Pontryagin's Maximum Principle. Then, using the optimal disturbances, we derive the robust filter recursions in Proposition 2. The robust filter safeguards against the misspecification risk inherent in making distributional assumptions. Furthermore, it reveals two new insights, i.e., (i) the robust gain matrix *optimally* enhances the weight on observed data compared to the Kalman filter, and (ii) the robust filter nests the Kalman filter as $\gamma \to \infty$. In other words, the Kalman filter is the *least* conservative in the family of conservative filters indexed by γ.

Applying the proposed method in the context of blood donation marketing, we find that the estimated effectiveness of marketing activities increases and the carryover effect decreases as conservatism increases. This compensatory effect generalizes across the five English-speaking Canadian provinces. We also uncover the novel finding that personal (i.e., phone call) communications are four times more effective than mass (i.e., radio) advertising. In Proposition 3, we further prove that nonprofit organizations should optimally advertise more (less) when sales are low (high), which is the *opposite* of the advertising-to-sales ratio heuristic used by for-profit companies. Finally, we find that robustification improves out-of-sample predictions and induces a profit-volatility trade-off.

We close by noting how to use the robust dynamic estimation. Operationally, users should make two small changes: (1) Replace the Kalman gain matrix with the new robust gain matrix in (17); and (2) Replace the prior covariance matrix with the new robust covariance matrix in (18). Specifically, replace the usual Kalman gain matrix $\tilde{K}_t = P_t^- Z_t'(Z_t P_t^- Z_t' + H_t)^{-1}$ with the robust gain matrix $K_t = P_t(I - P_t/\gamma + Z_t'H_t^{-1}Z_t P_t)^{-1}Z_t'H_t^{-1}$. Similarly, replace the usual prior covariance in period $(t+1)$ with $P_{t+1} = T_t P_t(I - P_t/\gamma + Z_t'H_t^{-1}Z_t P_t)^{-1}T_t' + Q_t$. Both the robust gain and covariance matrices involve subtraction of the matrix (P_t/γ), which is computationally simple. This small, but important, change yields the robust filter in Panel C. It optimally inflates the gain and covariance matrices (since the subtraction is inside an inverse operation) and places a greater weight on the market data (i.e., induces present-orientedness), thus accommodating more scenarios to consider when making decisions. Users can then perform the tests T_1 and T_2 to decide which method and results to use in their empirical contexts.

Endnote

[1] We thank the area editor for this suggestion.

References

Albers S, Mantrala MK, Sridhar S (2010) Personal selling elasticities: A meta-analysis. *J. Marketing Res.* 47(5):840–853.

Aravindakshan A, Rubel O, Rutz O (2014) Managing blood donations with marketing. *Marketing Sci.* 34(2):269–280.

Ataman MB, Mela CF, van Heerde HJ (2008) Building brands. *Marketing Sci.* 27(6):1036–1054.

Banavar RN (1992) Linear dynamic estimation. Unpublished doctoral dissertation, University of Texas, Austin.

Başar T, Bernhard P (1995) *H∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach* (Birkhäuser, Boston).

Biyalogorsky E, Naik PA (2003) Clicks and mortar: The effect of online activities and offline sales. *Marketing Lett.* 14(1):21–32.

Brockwell PJ, Davis RA (1996) *Introduction to Time Series and Forecasting* (Springer, New York).

Brodheim E, Prastacos GP (1979) The Long Island blood distribution system as a prototype for provincial blood management. *Interfaces* 9(5):3–20.

Bruce N (2008) Pooling and dynamic forgetting effects in multi-theme advertising: Tracking the advertising sales relationship with particle filters. *Marketing Sci.* 27(4):659–673.

Bruce NI, Peters K, Naik PA (2012) Discovering how advertising grows sales and builds brands. *J. Marketing Res.* 49(6):793–806.

Corstjens M, Merrihue J (2003) Optimal marketing. *Harvard Bus. Rev.* 81(10):114–121.

Eicker F (1963) Asymptotic normality and consistency of the least squares estimator for families of linear regression. *Ann. Math. Statist.* 34(2):447–456.

Eicker F (1967) Limit theorems for regression with unequal and dependent errors. Le Cam LM, Neyman J, eds. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* (University of California Press, Berkeley, CA), 59–82.

Farris P, Shames ER, Reibstein DJ (1998) *Advertising Budgeting: A Report from the Field* (American Association of Advertising Agencies, New York).

Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. A* 222:309–368.

Gatignon H (1993) Marketing mix models. Eliashberg J, Lilien GL, eds. *Handbooks in Operations Research and Management Science, Marketing,* Vol. 5 (North-Holland, Amsterdam), 697–732.

Gupta S, Steenburgh TJ (2008) Allocating marketing resources. Kerin RA, O'Regan R, eds. *Marketing Mix Decisions: New Perspectives and Practices* (American Marketing Association Publication, Chicago), 3–37.

Harvey AC (1989) *Forecasting Structural Time Series Models and the Kalman Filter* (Cambridge Universtity Press, Cambridge, UK).

Hinkley DV (1977) Jacknifing in unbalanced situations. *Technometrics* 19(3):285–292.

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. Le Cam LM, Neyman J, eds. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* (University of California Press, Berkeley, CA), 221–233.

Kalman RE (1960) A new approach to linear filtering and prediction problems. *J. Basic Engrg.* 82(March):35–45.

Kolsarici C, Vakratsas D (2015) Correcting for misspecification in parameter dynamics to improve forecast accuracy with adaptively estimated models. *Management Sci.* 61(10):2495–2513.

Koop G, Korobilis D (2010) Bayesian multivariate time series methods for empirical macroeconomics. *Foundations Trends Econometrics* 3(4):267–358.

Krever H (1998) Commission of Inquiry on the Blood System in Canada. Minister of Supply and Services, Ottawa, Ontario, Canada. http://publications.gc.ca/collections/Collection/CP32-62-3-1997-1E.pdf.

Kumar V, Mirchandani R (2012) Increasing the ROI of social media marketing. *MIT Sloan Rev.* 54(1):55–61.

Lachaab M, Ansari A, Jedidi K, Trabelsi A (2006) Modeling preference evolution in discrete choice models: A Bayesian state-space approach. *Quant. Marketing Econom.* 4(1):57–81.

Little JDC (1975) BRANDAID: A marketing-mix model, Part 1: Structure and Part 2: Implementation, calibration and case study. *Oper. Res.* 23(4):628–673.

Liu Y, Shankar V (2015) The dynamic impact of product-harm crises on brand equity and advertising effectiveness: An empirical analysis of the automobile industry. *Management Sci.* 61(10):2514–2535.

Lütkepohl H (2005) *New Introduction to Multiple Time Series Analysis* (Springer, Berlin).

Mantrala M (2002) Allocating marketing resources. Weitz BA, Wesley R, eds. *Handbook of Marketing* (Sage, London), 409–435.

Naik PA, Mantrala MK, Sawyer AG (1998) Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* 17(3):214–235.

Naik PA, Prasad A, Sethi SP (2008) Building brand awareness in dynamic oligopoly markets. *Management Sci.* 54(1):129–138.

Nichols W (2013) Advertising analytics 2.0. *Harvard Bus. Rev.* 91(3):60–68.

Prastacos G (1984) Blood inventory management: An overview of theory and practice. *Management Sci.* 30(7):777–800.

Raudenbush SW, Bryk AS (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods,* 2nd ed. (Sage, Thousand Oaks, CA).

Rubel O, Naik PA, Srinivasan S (2011) Optimal advertising when envisioning a product-harm crisis. *Marketing Sci.* 40(6): 1048–1065.

Sethi SP, Thompson GI (2000) *Optimal Control Theory: Applications to Management Science and Economics* (Springer, New York).

Sethuraman R, Tellis GT, Briesch RA (2011) How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *J. Marketing Res.* 48(3):457–471.

Shankar V (2008) Strategic allocation of marketing resources: Methods and insights. Kerin RA, O'Regan R, eds. *Marketing Mix Decisions: New Perspectives and Practices* (American Marketing Association, Chicago), 154–183.

Shumway RH, Stoffer DS (2011) *Time Series Analysis and Its Applications. With R Examples* (Springer-Verlag, New York).

Simon D (2006) *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches* (John Wiley and Sons, Hoboken, NJ).

Theodor Y, Shaked U (1994) Game theory approach to H∞-optimal discrete-time fixed-point and fixed-lag smoothing. *IEEE Trans. Automatic Control* 30(9):1944–1948.

Vakratsas D, Feinberg FM, Bass FM, Kalyanaram G (2004) The shape of advertising response function revisited: A model of dynamic probabilistic thresholds. *Marketing Sci.* 23(1):109–119.

van Heerde HP, Gijsenberg GJ, Dekimple MG, Steenkamp JB (2013) Price and advertising effectiveness over the business cycle. *J. Marketing Res.* 50(2):177–193.

Varian H (1975) A Bayesian approach to real estate assessment. Feinberg SE, Zellner A, eds. *Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage* (North-Holland, Amsterdam), 195–208.

Wald A (1950) *Statistical Decision Functions* (John Wiley, New York).

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

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4):817–838.

White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50(1):1–25.

Zames G (1981) Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Trans. Automatic Control* 26(2):301–320.

Zellner A (1986) Bayesian estimation and prediction using asymmetric loss functions. *J. Amer. Statist. Assoc.* 81(394):446–451.