SEMI–BLIND ROBUST IDENTIFICATION/MODEL (IN)VALIDATION WITH APPLICATIONS TO MACRO–ECONOMIC MODELLING.

W. Ma * **M. Yilmaz** * **M. Sznaier** * **C. Lagoa** *

* *Department of Electrical Engineering*
The Pennsylvania State University
University Park, PA 16802

Abstract: This paper addresses the problems of worst–case identification and model invalidation of systems subject to unknown initial conditions. While in principle these problems lead to non–convex Bilinear Matrix Inequalities (BMIs), we show that tractable convex relaxations are readily available. The potential of these techniques is illustrated by identifying and validating a subsystem of the macro–economy relating inflation and interest–rates.

Keywords: Robust Identification, Model (In)Validation, Macro-Economic Modelling

1. INTRODUCTION

Robust control tools have the potential to optimize the performance of economical and financial systems. Central to this approach is the modelling of the system under consideration as a linear time invariant (LTI) system subject to norm bound uncertainty. Robust optimization techniques can then be used to optimize the worst–case performance of the model (Sargent 1999, Tornell 2000, Sargent 2000, Stock 2000).

However, success of these techniques hinges upon the ability to obtain both a nominal model and an uncertainty description suitable to be used in a robust optimization context. Classical Bayesian analysis has been used extensively in economical systems modelling. In this approach, a parametric model, based on first principles is postulated and *a priori* probability distributions are assigned to each parameter. However, obtaining these parametric models and the associated probability distributions poses difficult practical problems (Rudebusch 1998). In addition, these probability distributions should be validated, as more experimental data becomes available. Wieland (1996) proposes to accomplish this by incorporating learning to update posteriors. However, since this approach is stochastic in nature, it can neither provide the uncertainty bounds required by the robust control tools nor conclusively invalidate the *a priori* assumptions.

To avoid these difficulties in this paper we propose to use robust identification and model (in)validation tools to obtain and validate models of economic subsystems, together with a worst–case uncertainty description. Accomplishing this requires extending currently existing techniques to handle unknown initial conditions. In principle this leads to an NP–hard problem involving bilinear matrix inequalities (BMIs). However, as we show in the paper, a convex relaxation is readily available.

The potential of the proposed approach is illustrated by using historical data from 21 quarters (starting with the first quarter of 1961, henceforth denoted as 1961.1, and ending in 1966.1) to identify and validate a model relating U. S. Federal Reserve Bank (FED) short-term interest rates (the input variable), to inflation rate (output variable). The resulting model was able to correctly forecast the inflation rate for the next 149 quarters (1966.2 through 2003.2). For comparison, the widely used Rudebusch–Svensson model (Rudebusch and Svensson 1998), obtained through a least squares fitting of the historical data from 1961.1 to 1996.2, and with the same order, yields a poorer fit.

2. PRELIMINARIES

2.1 Notation

In the sequel ℓ_p^m denotes the Banach space of vector-valued sequences equipped with the p norm. $\mathcal{L}_\infty$ denotes the Lebesgue space of complex-valued matrix functions essentially bounded on the unit circle, equipped with the norm $\|G\|_\infty \doteq \sup_{|z|=1} \overline{\sigma}(G(z))$; $\mathcal{H}_\infty$ the subspace of functions in $\mathcal{L}_\infty$ with bounded analytic continuation inside the unit disk, equipped with the norm $\|G\|_\infty \doteq \sup_{|z|<1} \overline{\sigma}(G(z))$; and $\mathcal{H}_{\infty,\rho}$ the space of transfer matrices in $\mathcal{H}_\infty$ equipped with the norm $\|G\|_{\infty,\rho} \doteq \sup_{|z|<\rho} \overline{\sigma}(G(z))$. Finally, $\mathcal{BX}(\gamma)$ denotes the open γ-ball in a normed space $\mathcal{X}$, and $\mathcal{BX}$ the open unit ball in $\mathcal{X}$.

From an input–output viewpoint any operator of interest G will be represented either by a (rational) complex–valued transfer function: $G(z) \doteq \sum_{k=0}^{\infty} g_k z^k$ or a minimal state–space realization:

$$G \equiv \left(\begin{array}{c|c} \mathsf{A} & \mathsf{B} \\ \hline \mathsf{C} & \mathsf{D} \end{array}\right) \tag{1}$$

In the sequel we will denote by $T_G : \ell^\infty[0,\infty) \to \ell^\infty[0,\infty)$ and $\Gamma_G : \ell^\infty(-\infty,-1] \to \ell^\infty[0,\infty)$ the Toeplitz and Hankel operators associated with an ℓ^∞ stable system G. Further, when dealing with finite sequences of length N we will represent these operator by the finite matrices T_G^N:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} g_0 & 0 & 0 & \dots \\ g_1 & g_0 & 0 & \dots \\ g_2 & g_1 & g_0 & \\ \vdots & \vdots & \ddots & \vdots \\ g_{N-1} & g_{N-2} & \cdots & g_o \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N-1} \end{bmatrix}. \tag{2}$$

and H_G^N:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} g_1 & g_2 & \cdots & g_N \\ g_2 & g_3 & \cdots & g_{N+1} \\ \vdots & \vdots & \dots & \vdots \\ g_N & g_{N+1} & \cdots & g_{2N-1} \end{bmatrix} \begin{bmatrix} u_{-1} \\ u_{-2} \\ u_{-3} \\ \vdots \\ u_{-N} \end{bmatrix}. \tag{3}$$

2.2 Background results on interpolation theory

The following results will be used in the paper to establish the existence of LTI systems with the appropriate features.

Lemma 1. (Carathéodory-Fejér). Given a matrix valued sequence $\{\mathsf{L}_i\}_{i=0}^{n-1}$, there exists a causal, discrete-time, LTI operator $L(z) \in \mathcal{BH}_\infty$ such that

$$L(z) = \mathsf{L}_0 + \mathsf{L}z + \mathsf{L}_2 z^2 + \dots \mathsf{L}_{n-1} z^{n-1} + \dots \tag{4}$$

if and only if $(\mathsf{T}_L^n)^T \mathsf{T}_L^n \leq \mathsf{I}$.

Proof: See for instance Foias and Frazho (1990).

In the sequel we will consider systems of the form $\mathcal{S}$:

$$\mathcal{S} \doteq \{G(z) = H(z) + P(z)\}. \tag{5}$$

where $H(z) \in \mathcal{BH}_{\infty,\rho}(K)$ for some $\rho \geq 1$ and $P(z)$ represent the nonparametric and parametric components of the operator respectively. We will further assume that $P(z)$ belongs to the following class $\mathcal{P}$ of affine operators:

$$\mathcal{P} \doteq \{P(z) = \mathbf{p}^T \mathbf{G}_p(z),\ \mathbf{p} \in \mathcal{R}^{N_p}\}, \tag{6}$$

where the N_p components $\mathbf{G}_{p_i}(z)$ of vector $\mathbf{G}_p(z)$ are known, linearly independent, rational transfer functions. The following result gives a necessary and sufficient condition for two finite vector sequences to be related by an operator in the family $\mathcal{S}$.

Lemma 2. [Parrilo *et al.* (1999)] Given K, ρ and two vector sequences $(\mathbf{u}, \mathbf{y})$, there exists an operator $S \in \mathcal{S}$ such that $\mathbf{y} = S\mathbf{u}$ if and only if there exists a vector $\mathbf{h}$ satisfying:

$$\begin{gathered} M(\mathbf{h}) \doteq \begin{bmatrix} K\mathsf{R}^{-2} & (\mathsf{T}_h^N)^T \\ \mathsf{T}_h^N & K\mathsf{R}^2 \end{bmatrix} \geq 0 \\ \mathbf{y} = \mathsf{T}_u^N \mathsf{P}\mathbf{p} + \mathsf{T}_u^N \mathbf{h} \end{gathered} \tag{7}$$

where $(\mathsf{P})_k \doteq [g_k^1\ g_k^2\ \cdots\ g_k^{N_p}]$,where g_k^i denotes the k-th Markov parameter of the i-th transfer function $G_{p_i}(z)$, h_k is the k-th Markov parameter of the nonparametric component $H(z)$, T_u^N is the lower Toepliz matrix associated with the sequence $\mathbf{u}$, and where

$$R = \text{diag}\left[\ 1\ \rho\ \rho^2\ \dots \rho^{N-1}\ \right]$$

Moreover, in this case all such operators S can be parameterized in in terms of a free parameter $Q(z) \in \mathcal{BH}_\infty$. In particular, the choice $Q(z) = 0$ leads to the "central" model $S_{\text{central}}(z) = H_o(z) + \mathbf{p}^T \mathbf{G}_p(z)$ where an explicit state–space realization of $H_o(z)$ can be found for instance in Parrilo *et al.* (1998)

3. SEMI–BLIND IDENTIFICATION

3.1 Motivation: Macro-Economic Modelling

The goal of this paper is to illustrate the applicability of robust identification/model (in)validation to macro-economic modelling. To this effect, we will obtain and validate a model of the subsystem of the US economy relating FED funds rates to inflation. An empirical, simplified model of this subsystem, the Rudebusch and Svensson(RS) model, is already available in the literature (Rudebusch and Svensson 1998):

$$\begin{aligned} \pi_{t+1} &= 0.7\pi_t - 0.1\pi_{t-1} + 0.28\pi_{t-2} + \\ &+ 0.12\pi_{t-3} + 0.14y_t + \epsilon_{t+1} \\ y_{t+1} &= 1.16y_t - 0.25y_{t-1} - 0.1(\overline{i_t} - \overline{\pi_t}) \\ &+ \eta_{t+1} \end{aligned} \tag{8}$$

where $\pi_t = 400.[ln(p_t) - ln(p_{t-1})]$ is the quarterly inflation in the Gross-Domestic Product(GDP) chain-weighted price index(p_t) in percent at an annual rate,

$\overline{\pi_t} = 0.25.\sum_{k=0}^{3}(\pi_{t-k})$ is the 4-quarter average of the inflation, $y_t = 100(\frac{q_t - q_t^*}{q_t^*})$ is the percentage gap between actual real(q_t) and potential GDP(q_t^*), i_t is the quarterly Fed funds rate in percent at an annual rate, $\overline{i_t} = 0.25.\sum_{k=0}^{3}(i_{t-k})$ is the 4-quarter average of the Fed funds rate and ϵ_{t+1}, η_{t+1} are i.i.d zero mean disturbances. This model was obtained by least squares fitting of historical data from the first quarter of 1961 through the second quarter of 1996.

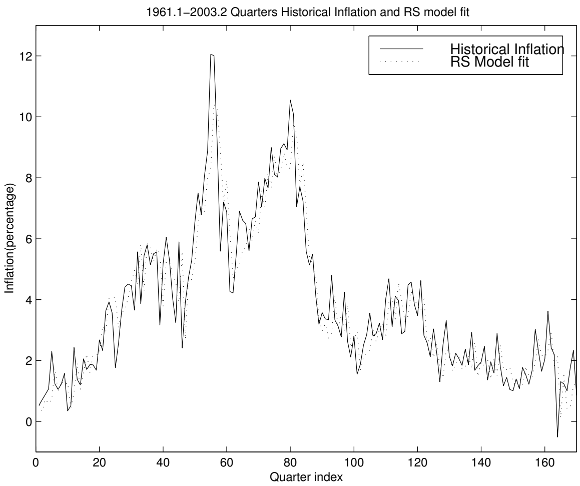

Fig. 1. Historical Inflation Data and RS model fit, maximum error is 3.75

As shown in Figure 1, the RS model indeed fits the training data. Note however that (i) this technique requires a large amount of data points, while at the same time not exploiting *a priori* information available about the subsystem in question, and, (ii) no conclusions can be drawn about modelling and future prediction errors. The goal of this paper is to show that these difficulties can be avoided by using a deterministic, set membership approach. Robust identification of LTI systems has been well studied during the past few years (see for instance the textbook (Chen and Gu 2000). In particular, the method that we will use in this paper is an extension of the mixed parametric/non–parametric approach introduced by Parrilo *et al.* (1999). However, contrary to the case commonly addressed in robust identification, only a partial (in this case post 1961.1) experimental data record is available. The effect of inputs prior to 1961.1 are encapsulated in some (unknown) non zero initial conditions x_o. Ignoring these initial conditions would lead to artificially high identification errors.

3.2 Identification with unknown initial conditions

The presence of unknown, non–zero initial conditions noted above motivates the following *semi-blind* identification problem:

Problem 1. Given an unknown plant, *a priori* sets of candidate models and noise $(\mathcal{S},\mathcal{N})$ and a finite set of samples of the input $\mathbf{u}$ to the plant and its corresponding output $\mathbf{y}$ corrupted by additive measurement noise η, find a model g compatible with both the *a priori* information and the *a posteriori* experimental data, that is $g \in \mathcal{T}(\mathbf{y})$, where

$$\mathcal{T}(\mathbf{y}) \doteq \{g \in \mathcal{S} : y_k - \sum_{i=0}^{k} h_i u_{k-i}$$
$$-C_g{*}A_g^{k-1}\mathbf{x}_o \in \mathcal{N},\ k = 0,\dots,N-1\},\ \text{for some } \mathbf{x}_o \tag{9}$$

where

$$g = \left(\begin{array}{c|c} A_g & B_g \\ \hline C_g & D_g \end{array}\right),\ h_o = D_g,\ h_i = C_g A_g^{i-1} B_g$$

It can be shown, by invoking Lemma 2, that the problem above reduces to a BMI in $\{h_i\}$ and $\mathbf{x}_o$. However, BMIs generically lead to non–convex, NP–hard optimization problems. To avoid this difficulty, in the sequel we propose a convex relaxation of Problem 1. To this effect, we will assume that the *past* inputs to the system are known to belong to some set $\mathcal{U}_-$ [1]. Thus the effect of the unknown initial condition $\mathbf{x}_o$ can be replaced by the effect of an unknown signal $u_- \in \mathcal{U}_-$ acting in $(-\infty,-1]$. This leads to the following reformulation of equation (9):

$$\begin{aligned}\mathcal{T}(\mathbf{y}) \doteq \{g \in \mathcal{S} : y_k - (\Gamma_g u_-)_k - (T_g u_+)_k \in \mathcal{N} \\ k = 0,\dots,N-1\},\ \text{for some } u_- \in \mathcal{U}_-\end{aligned} \tag{10}$$

where $u_+ \doteq \{u_0, u_1, \dots, u_{N-1}\}$ and where Γ_g and T_g represent the Hankel and Toeplitz operators associated with the system g respectively. Next, replace $(\Gamma_g u_-)_k$ by a new variable x_k subject to the constraint that $x_k = (\Gamma_g u_-)_k$ for some $u_- \in \mathcal{U}_-$. Assuming that the set of admissible past inputs has the form $\mathcal{U} = \mathcal{B}\ell^p(K_u)$ and that a bound $\|\Gamma_g\|_{\ell^p \to \ell^\infty} \leq \gamma$ is available as part of the *a priori* information, this leads to the following convex relaxation of problem 1:

Problem 2. Given an unknown plant, the *a priori* sets of candidate models, past inputs and noise $(\mathcal{S},\mathcal{U}_-,\mathcal{N})$ and a finite set of samples of the input and output of the plant $(\mathbf{u},\mathbf{y})$ in $[0, N-1]$, find a model $g \in \mathcal{T}(\mathbf{y})$, where:

$$\begin{aligned}\mathcal{T}(\mathbf{y}) \doteq \{g \in \mathcal{S} : y_k - x_k - (T_g u_+)_k \in \mathcal{N} \\ \text{for some } |x|_k \leq \gamma K_u, k = 0,\dots,N-1\}\end{aligned} \tag{11}$$

Straightforward application of Lemma 2 leads now to the following result:

Proposition 1. Problem 2 has a solution if and only if the following set of LMIs in $\mathbf{h},\mathbf{x}$ is feasible:

[1] As we will show in the sequel, this assumption holds for the economic subsystem under consideration here.

$$
\begin{aligned}
M(\mathbf{h}) \doteq \begin{bmatrix} K\mathsf{R}^{-2} & (\mathsf{T}_h^N)^T \\ \mathsf{T}_h^N & K\mathsf{R}^2 \end{bmatrix} \geq 0 \\
\mathbf{y} - \mathsf{T}_u^N \mathsf{P}\mathbf{p} - \mathsf{T}_u^N \mathbf{h} - \mathbf{x} \in \mathcal{N} \\
-\gamma K_u \leq \mathbf{x} \leq \gamma K_u
\end{aligned} \tag{12}
$$

where the last two inequalities should be interpreted in a component–wise sense.

3.3 Identification results:

The experimental data used in this paper, obtained from the websites `www.marketvector.com` and `www.federalreserve.gov`, consists of historical values of the quarterly U.S. FED funds rate, i_k, and the corresponding values of the inflation y_k. In order to account for the difficulty in exactly measuring inflation, in the sequel we will assume that $y_k = \pi_k + \eta_k$, where y_k, π_k, η_k denote the measured inflation, the actual inflation and additive noise, respectively. We will postulate a simplified scenario where the inflation π is assumed to be the output of an unknown, stable LTI system in response to the input sequence i_k. In order to apply the framework discussed in the previous section, we need a characterization of the set of past inputs $\mathcal{U}_-$. In principle, one could just characterize this set as $0 \leq i_- \leq i_{imax}$ where i_{max} is some bound on the maximum historical FED funds rate. However, this bound is potentially too coarse, since the FED funds rate has changed considerably over the period under consideration. On the other hand, the quarterly *change* in the FED funds rate is substantially lower and more uniform across the period of interest. Thus, a much tighter bound on the past inputs can be obtained by identifying an operator S mapping the *change* in FED funds rate, $u_k = i_k - i_{k-1}$, to inflation π_{k+1}. Note that this operator should then include an integrator in its parametric portion. Finally, from a Fourier analysis of the input/output data it was determined that an upper bound of ρ is given by $\rho \leq 1.02$. With these assumptions, the *a priori* information is given by:

$$
\begin{aligned}
\mathcal{S} &= \Big\{ H(z) = p_1 \frac{z}{z-1} + G_{np}(z) \,| \\
&\qquad G_{np}(z) \in \mathcal{BH}_{\infty,\rho}(K) \Big\} \\
\mathcal{N} &= \{\eta \in \ell^\infty \colon |\eta_k| \leq \epsilon\}\,,\ \epsilon = 0.2 \\
\mathcal{U}_- &= \{u \in \ell^\infty \colon |u_k| \leq u_{max}\}\,, u_{max} = 0.5 \\
x_k &= p \cdot i_k + x_k^{np},\ |x_n^{np}| \leq K \frac{\rho^{-n}}{\rho - 1} 0.5
\end{aligned} \tag{13}
$$

where the last equation comes from the facts that:

$$
x_k = p \sum_{j=-\infty}^{k} (i_j - i_{j-1}) + \Gamma_{np} u_- = i_k + \Gamma_{np} u_-
$$

$$
|(\Gamma_{np} u_-)_k| = |\sum_{j=-\infty}^{-1} g_{k-j}^{np} u_j| \leq |u_{max}| \sum_{k+1}^{\infty} |g_j^{np}|
$$

and $G^{np} \in \mathcal{BH}_{\infty,\rho}(K) \Rightarrow |g_n^{np}| \leq K\rho^{-n}$.

Running the identification algorithm outlined in the previous section using $N = 21$ historical values of inflation and FED funds rates (from 1961.1 to 1966.1) led (after eliminating unobservable/uncontrollable states) to the following 4^{th} order model:

$$
\begin{aligned}
\pi_{k+4} &= 0.526\pi_{k+3} + 0.074\pi_{k+2} - 0.13\pi_{k+1} \\
&+ 0.506\pi_k + 0.522u_{k+4} + 0.208u_{k+3} \\
&+ 0.184u_{k+2} + 0.276u_{k+1} - 0.024u_k
\end{aligned} \tag{14}
$$

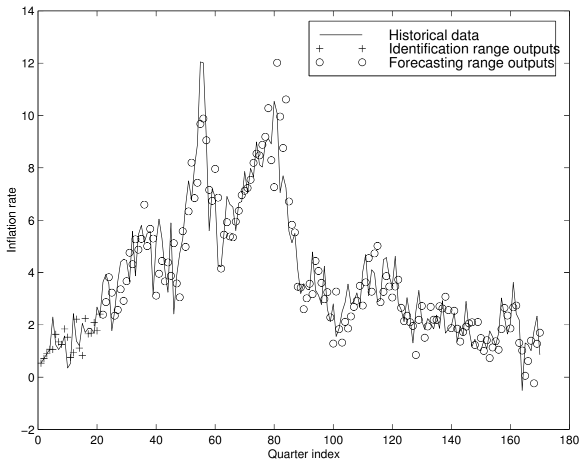

Fig. 2. Historical (solid line) versus predicted ('o') inflation: '+' indicates a point used in the identification

The forecasting power of this model is illustrated in Figure 2, where it was used to estimate the inflation for the entire 1961.1-2003.2 quarters, using as inputs past historical data for inflation and changes in FED funds rates. As shown there, the model (14) yields a worst–case error of 3.37. For comparison, the RS model has a worst case error of 3.75, (roughly 11% higher) even though it was obtained considering 142 data points.

3.4 Worst case prediction error bounds

As we illustrate next, one of the advantages of the approach outlined above, is its ability to provide worst–case bounds on the prediction error. Begin by noting that the algorithm is interpolatory, that is it produces a model inside the consistency set $\mathcal{T}(\mathbf{y})$. Thus, since the "true" system must also belong to the consistency set [2], it follows that, given the first N measurements $y_i,\ i = 0, \ldots, N-1$ a bound on the worst case prediction error at $t = N$ is given by:

$$
\begin{aligned}
|e_N| \leq \sup_{g_1, g_2 \in \mathcal{T}(\mathbf{y})} |\,[(T_{g_1} - T_{g_2})\mathbf{u}_+ + \Gamma_{g_1}\mathbf{u}_{-1} \\
- \Gamma_{g_2}\mathbf{u}_{-2}]_N\,| = d[\mathcal{T}(\mathbf{y})] \leq \sup_{\mathbf{y}} d[\mathcal{T}(\mathbf{y})] = \mathcal{D}(\mathcal{I})
\end{aligned} \tag{15}
$$

where $d(.)$ and $\mathcal{D}(\mathcal{I})$ denote the diameter of the set $\mathcal{T}(\mathbf{y})$, in the sup–metric, and the diameter of information, respectively. Moreover, since the *a priori* sets $(\mathcal{S}, \mathcal{N})$ are convex and symmetric, with points of symmetry $g_s = 0$ and $\eta_s = 0$ respectively, it can be shown (see for instance Lemma 10.3 in (Sánchez Peña and Sznaier 1998)) that:

$$
\mathcal{D}(\mathcal{I}) \leq 2 \sup_{g \in \mathcal{T}(\mathbf{0})} \left| p \cdot i_N + \sum_{j=0}^{N} h_{N-j} \cdot u_j + K \cdot K_u \frac{\rho^{-N+1}}{\rho - 1} \right| \tag{16}
$$

[2] As long as the *a priori* information is indeed correct.

where $\mathcal{T}(\mathbf{0})$ indicate the set of operators compatible with the zero outcome: $y_k = 0,\ k = 0, 1, \ldots, N-1$. This leads to the following convex optimization problem:

$$\max |p \cdot i_N + \sum_{j=0}^{N} h_{N-j} u_j + K \cdot K_u \frac{\rho^{-N+1}}{\rho - 1}| \quad (17)$$

subject to:

$$\begin{aligned} &M(\mathbf{h}) \doteq \begin{bmatrix} K\mathsf{R}^{-2} & (\mathsf{T}_h^N)^T \\ \mathsf{T}_h^N & K\mathsf{R}^2 \end{bmatrix} \geq 0 \\ &\|\mathsf{T}_u \mathsf{P}\mathbf{p} + \mathsf{T}_u \mathbf{h} + \mathbf{x}\|_{\ell^\infty} \leq \epsilon \\ &|x_j| \leq K \cdot K_u \frac{\rho^{-j}}{\rho - 1},\ |u(N)| \leq u_{max} \\ &|i_N| \leq i_{max}, |h_N| \leq K\rho^{-N} \end{aligned} \quad (18)$$

Next we illustrate the use of this bound by obtaining an estimate of the prediction error at $N = 22$, the first historical data point not used in the identification. Solving the problem above with $u_{max} = 0.6$ and $i_{max} = 6$, yields $|e_{22}| \leq 2.37$ (actual error is 1.05). The bounds corresponding to $N = 25$ and $N = 30$ are 1.99 and 1.95, showing that, as expected, the error gets smaller as more points are used in the identification.

4. SEMI–BLIND MODEL (IN)VALIDATION

Next we turn our attention to the related problem of model (in)validation. From a practical standpoint, before using the model (14) to decide macro-economic policy, it should be validated using *new* data, that has not been used in the identification process (to avoid introducing biases). In addition, this process will provide worst–case bounds on the model uncertainty associated with the description (14) [3] Finally, this validation step can also indicate when the model is no longer compatible with the measured data, for instance due to changing parameters, providing a mechanism to answer Lucas' critique (Lucas 1976) questioning the usefulness of using identified models for macroeconomic forecasting.

Assuming multiplicative uncertainty and additive noise leads to the following (in)validation problem:

Problem 3. Given experimental data $\{\mathbf{y}_k, \mathbf{u}_k^+\}$, consisting of N measurements of the quarterly inflation and the change in Federal Reserve funds rate, a nominal model S and set descriptions $\mathcal{N}$,$\mathbf{\Delta}$ and $\mathcal{X}_o$ of admissible noise, uncertainty and initial conditions, determine if there exists at least one triple $(\eta, \Delta, \mathbf{x}_o) \in \mathcal{N} \times \mathbf{\Delta} \times \mathcal{X}_o$ that can reproduce the available experimental evidence:

$$\mathbf{y} = (I + \Delta)(T_S \mathbf{u}^+ + T_S^{ic} \mathbf{x}_o) + \eta \qquad (19)$$

where T_S and T_S^{ic} denote the operators that map the input and initial conditions of system S to its output.

[3] these bounds are typically substantially tighter than the ones obtained by computing an upper bound on the diameter of information (Chen and Gu 2000).

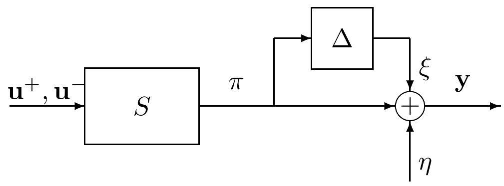

Fig. 3. The Set-Up for Semi–Blind Model Invalidation.

Compared to standard invalidation problems, the problem above has an additional term due to unknown initial conditions. As in the identification case, this term can be replaced by the action of some unknown input acting in $[-T, -1]$, for some $T > 0$, that is, establishing existence of a triple $(\eta, \Delta, \mathbf{u}^-) \in \mathcal{N} \times \mathbf{\Delta} \times \mathcal{U}_-$ such that:

$$\mathbf{y} = (I + \Delta)(T_S \mathbf{u}^+ + \Gamma_S \mathbf{u}^-) + \eta \qquad (20)$$

where $\mathcal{U}_-$ and Γ_S denote the set of admissible past inputs and the Hankel operator associated with S, respectively. Equivalently (see Figure 3), the model is not invalidated by the experimental data if and only if there exists some $\Delta \in \mathbf{\Delta}$ such that $\mathbf{y} - \pi - \eta \doteq \xi = \Delta\pi$. Straightforward application of Lemma 1 assuming unstructured uncertainty of the form $\mathbf{\Delta} = \mathcal{BH}_\infty(\delta)$ shows that this is equivalent to feasibility of the following inequalities

$$\begin{aligned} &(\mathsf{T}_\xi^N)^T (\mathsf{T}_\xi^N) \leq \delta^2 (\mathsf{T}_\pi^N)^T (\mathsf{T}_\pi^N) \iff \delta^2 (\mathsf{T}_\pi^N)^T (\mathsf{T}_\pi^N) \\ &- (\mathsf{T}_y^N - \mathsf{T}_\pi^N - \mathsf{T}_\eta^N)^T (\mathsf{T}_y^N - \mathsf{T}_\pi^N - \mathsf{T}_\eta^N) \geq 0 \end{aligned} \qquad (21)$$

where $\mathsf{T}_\pi^N \doteq \mathsf{T}_S^N \mathsf{T}_{u^+}^N + \mathsf{H}_S^{N-} \mathsf{T}_{u^-}^{N-}$, and where T_S^N and H_S^{N-} are the (finite) Toeplitz and Hankel matrix associated with the system S. Unfortunately (21) is not jointly convex on all the variables involved, due to the cross–terms ${\mathsf{T}_\eta^N}^T \mathsf{H}_S^{N-} \mathsf{T}_{u^-}^{N-}$. To avoid the difficulties associated with solving non–convex problems, in the sequel we propose a tractable convex relaxation.

4.1 A Deterministic Convex Relaxation

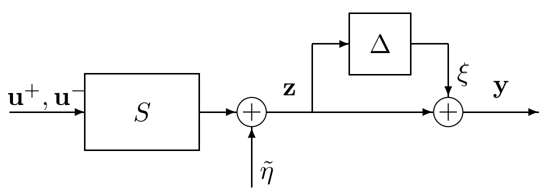

Fig. 4. A Convex Relaxation for Semi–Blind Invalidation.

Consider the alternative setup shown in Fig. 4, where the measurement noise is also affected by the unknown error dynamics Δ:

$$\mathbf{y} = (I + \Delta)(T_S \mathbf{u}^+ + \Gamma_S \mathbf{u}^- + \tilde{\eta}) \qquad (22)$$

When compared to the original setup shown in Fig. 3, it can be easily seen that the only difference is in the measurement noise level. Specifically, if there exists a triple $(\mathbf{u}^-, \tilde{\eta}, \Delta)$ satisfying (22) with $\|\tilde{\eta}\|_2 \leq \tilde{\epsilon}$ and

Table 1. Model Invalidation Results: Identified versus R.S. model

Data range	Prediction error	$\|\Delta\|_\infty$	$\|\Delta_{RS}\|_\infty$
5-170	1.7	0.5	0.63
5-170	2	0.3	0.4575

$\|\Delta\|_\infty \leq \delta$, then the triple $(\mathbf{u}^-, \eta, \Delta)$ with $\eta \doteq (1+\Delta)\tilde{\eta}$ satisfies (20). Thus, one can attempt to find a solution to the original problem by searching for a solution to the model (in)validation problem shown in Fig. 4, with noise level $\tilde{\epsilon} \doteq \frac{\epsilon}{1+\delta}$. As we show in the sequel this leads to a convex optimization problem. In addition, one will expect that if $\|\Delta\|_\infty \ll 1$ then this approximation is not too conservative. This conjecture will be experimentally substantiated in section 4.2.

Theorem 1. There exists a feasible triple $(\eta, \Delta, \mathbf{u}^-) \in \mathcal{N} \times \mathbf{\Delta} \times \mathcal{U}_-$ that satisfies (22) if and only if there exists at least one pair of finite sequences $\mathbf{u}_- = \{\mathbf{u}^-_{-1}, \mathbf{u}^-_{-2}, \cdots, \mathbf{u}^-_{-N_--1}\} \in \mathcal{U}_-$ and $\tilde{\eta} \in \tilde{\mathcal{N}}$ and some $\delta < 1$ such that the following LMI holds:

$$\mathsf{A}_1 \doteq \begin{bmatrix} \mathsf{X}(\mathbf{u}_-) & (\mathsf{T}^N_{\tilde{\omega}})^T \\ \mathsf{T}^N_{\tilde{\omega}} & (\delta^2-1)^{-1}\mathsf{I} \end{bmatrix} \leq \quad 0 \qquad (23)$$

where:

$$T^N_{\tilde{\omega}} \doteq \mathsf{T}^N_S \mathsf{T}^N_{u^+} + \mathsf{H}^{N_-}_S \mathsf{T}^{N_-}_{u^-}$$
$$\mathsf{X} \doteq (T^N_y)^T \cdot T^N_y - (T^N_y)^T \cdot T^N_{\tilde{\omega}} - (T_{\tilde{\omega}^N})^T \cdot T^N_y$$

Proof: (Sketch). The proof follows from applying Lemma 1 to the signals z and ξ (in Fig. 4) and straightforward manipulations (omitted for space reasons).

4.2 *Experimental Results:*

The (in)validation procedure outlined above was applied to the model (14) identified in section 3.3, seeking to establish the minimum size of the uncertainty such that, for a given measurement error level, the model is not invalidated by the experimental data. For comparison purposes the same process was performed using the empirical RS model. The results are summarized in Table 1. In all cases the initial conditions at $t = 0$ were assumed to have been reached by a sequence of past inputs $u_- \in [-4, -1])$ with $|u_k| \leq 0.5$.

5. CONCLUSIONS

Many problems of practical interest involve identifying and validating models from experimental data, but where the experiment cannot be controlled, in the sense that the researcher can choose neither the test input nor the initial conditions of the system under observation. This leads to non–convex, generically NP hard optimization problems. To avoid this difficulty, in this paper we propose tractable convex relaxations that perform well in practice. These results were illustrated by identifying and validating an economic model relating interest rates to inflation. The resulting model was shown to outperform (by as much as 50% as in the last row of Table 1) a widely used linear model, even though the latter was obtained using data registers 7 times longer. Research currently under way seeks to apply this technique to similar problems such as analysis and classification of human activities.

ACKNOWLEDGMENTS

This work was supported by NSF grants ECS–0115946 and ECS–0221562.

REFERENCES

Chen, J. and G. Gu (2000). *Control Oriented System Identification, An $\mathcal{H}_\infty$ Approach*. John Wiley. New York.

Foias, C. and A. E. Frazho (1990). *The commutant lifting approach to interpolation problems, Operator theory: Advances and Applications*. Vol. 44. Birkhäuser.

Lucas, Robert E., Jr. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Series on Public Policy* **1**, 19–46.

Parrilo, P. A., M. Sznaier and R. S. Sanchez Pena (1998). Mixed time/frequency domain based robust identification. *Automatica* **34**(11), 1375–1389.

Parrilo, P. A., R. S. Sanchez Pena and M. Sznaier (1999). A parametric extension of mixed time/frequency domain based robust identification. *IEEE Trans. Autom. Contr.* **44**(2), 364–369.

Rudebusch, G.D. (1998). Is the fed too timid? monetary policy in an uncertain world. Manuscript, Research Department, Federal Reserve Bank of San Francisco.

Rudebusch, Glenn and Lars Svensson (1998). Policy rules for inflation targeting. Working paper 6512, NBER.

Sánchez Peña, R. and M. Sznaier (1998). *Robust Systems Theory and Applications*. Wiley & Sons, Inc.

Sargent, J. Thomas (1999). Comment. In: *Monetary Policy Rules* (John B. Taylor, Ed.). pp. 144–154. University of Chicago Press. Chicago.

Sargent, J. Thomas (2000). Wanting robustness in macroeconomics and robust control and filtering of forward-looking models.

Stock, James (2000). Robust monetary policy under model uncertainty in a small model of the US economy.

Tornell, Aaron (2000). Robust-H_∞ forecasting and asset pricing anomalies. Working paper 7753, NBER.

Wieland, V. (1996). Monetary policy, parameter uncertainty and optimal learning. Manuscript, Federal Reserve Board.