# A large-scale application of Stata's forecast suite: challenges and potential

Christopher F Baum

*Boston College and DIW Berlin*

SUGUK 2015, London

# Stata's forecast suite

- A very welcome addition to Stata's capabilities for econometric modeling arrived in Version 13: the `forecast` suite of commands.
- These commands support the definition of an *econometric model:* a set of estimated equations and identities that can be solved for a set of *endogenous variables* as functions of the *exogenous variables*.
- Nonlinear relationships and transformations can be used. For instance, an estimated equation may express the percentage change in an endogenous variable, while the level of that variable or its logarithm may appear in other equations.

# Stata's forecast suite

- A very welcome addition to Stata's capabilities for econometric modeling arrived in Version 13: the `forecast` suite of commands.
- These commands support the definition of an *econometric model:* a set of estimated equations and identities that can be solved for a set of *endogenous variables* as functions of the *exogenous variables*.
- Nonlinear relationships and transformations can be used. For instance, an estimated equation may express the percentage change in an endogenous variable, while the level of that variable or its logarithm may appear in other equations.

# Stata's forecast suite

- A very welcome addition to Stata's capabilities for econometric modeling arrived in Version 13: the `forecast` suite of commands.
- These commands support the definition of an *econometric model:* a set of estimated equations and identities that can be solved for a set of *endogenous variables* as functions of the *exogenous variables*.
- Nonlinear relationships and transformations can be used. For instance, an estimated equation may express the percentage change in an endogenous variable, while the level of that variable or its logarithm may appear in other equations.

- This provides a feature not generally available to Stata's time series estimators: the ability to produce true *dynamic forecasts* of a single equation or of multiple equations.
- In a dynamic forecast, values of endogenous variables computed in one period are plugged in to the prediction equation in the next period. When the equation contains one or more lagged dependent variables, it is said to be dynamic. A one-period dynamic forecast from that equation is identical to a static forecast, but beyond one period ahead, they differ.
- Although `arima`, `var`, `svar`, `arch` and `mgarch` have postestimation commands that implement dynamic forecasts, such forecasts could not easily be computed after `regress`, `ivregress`, `sureg`, `reg3`, etc.
- Stata's `forecast` commands close this gap.

- This provides a feature not generally available to Stata's time series estimators: the ability to produce true *dynamic forecasts* of a single equation or of multiple equations.
- In a dynamic forecast, values of endogenous variables computed in one period are plugged in to the prediction equation in the next period. When the equation contains one or more lagged dependent variables, it is said to be dynamic. A one-period dynamic forecast from that equation is identical to a static forecast, but beyond one period ahead, they differ.
- Although `arima`, `var`, `svar`, `arch` and `mgarch` have postestimation commands that implement dynamic forecasts, such forecasts could not easily be computed after `regress`, `ivregress`, `sureg`, `reg3`, etc.
- Stata's `forecast` commands close this gap.

- This provides a feature not generally available to Stata's time series estimators: the ability to produce true *dynamic forecasts* of a single equation or of multiple equations.
- In a dynamic forecast, values of endogenous variables computed in one period are plugged in to the prediction equation in the next period. When the equation contains one or more lagged dependent variables, it is said to be dynamic. A one-period dynamic forecast from that equation is identical to a static forecast, but beyond one period ahead, they differ.
- Although `arima`, `var`, `svar`, `arch` and `mgarch` have postestimation commands that implement dynamic forecasts, such forecasts could not easily be computed after `regress`, `ivregress`, `sureg`, `reg3`, etc.
- Stata's `forecast` commands close this gap.

- This provides a feature not generally available to Stata's time series estimators: the ability to produce true *dynamic forecasts* of a single equation or of multiple equations.
- In a dynamic forecast, values of endogenous variables computed in one period are plugged in to the prediction equation in the next period. When the equation contains one or more lagged dependent variables, it is said to be dynamic. A one-period dynamic forecast from that equation is identical to a static forecast, but beyond one period ahead, they differ.
- Although `arima`, `var`, `svar`, `arch` and `mgarch` have postestimation commands that implement dynamic forecasts, such forecasts could not easily be computed after `regress`, `ivregress`, `sureg`, `reg3`, etc.
- Stata's `forecast` commands close this gap.

- The ability to define and solve a complex, nonlinear econometric model in Stata brings its capabilities in line with a number of other econometric software tools.
- In this talk, I will discuss my experience implementing a nontrivial econometric model using Stata's `forecast`, and highlight a number of issues that arose in this process.
- To be truly competitive with well-established econometric software, StataCorp should address these issues.

- The ability to define and solve a complex, nonlinear econometric model in Stata brings its capabilities in line with a number of other econometric software tools.
- In this talk, I will discuss my experience implementing a nontrivial econometric model using Stata's `forecast`, and highlight a number of issues that arose in this process.
- To be truly competitive with well-established econometric software, StataCorp should address these issues.

- The ability to define and solve a complex, nonlinear econometric model in Stata brings its capabilities in line with a number of other econometric software tools.
- In this talk, I will discuss my experience implementing a nontrivial econometric model using Stata's `forecast`, and highlight a number of issues that arose in this process.
- To be truly competitive with well-established econometric software, StataCorp should address these issues.

- To gain perspective on the problem: the Stata manual description of `forecast` illustrates its application to the venerable Klein Model I, a structure described in many econometrics textbooks.
- Klein Model I contains a total of three estimated equations and four identities: seven endogenous variables, which are functions of four exogenous variables.
- While useful as a textbook exercise, implementing this model does not illuminate the challenges faced when the `forecast` tools are used for a professional forecasting model of the US macroeconomy.
- I describe our motivation to implement such a model in work I have been doing with a major US financial institution, Citizens Financial Group (CFG).

- To gain perspective on the problem: the Stata manual description of `forecast` illustrates its application to the venerable Klein Model I, a structure described in many econometrics textbooks.
- Klein Model I contains a total of three estimated equations and four identities: seven endogenous variables, which are functions of four exogenous variables.
- While useful as a textbook exercise, implementing this model does not illuminate the challenges faced when the `forecast` tools are used for a professional forecasting model of the US macroeconomy.
- I describe our motivation to implement such a model in work I have been doing with a major US financial institution, Citizens Financial Group (CFG).

- To gain perspective on the problem: the Stata manual description of `forecast` illustrates its application to the venerable Klein Model I, a structure described in many econometrics textbooks.
- Klein Model I contains a total of three estimated equations and four identities: seven endogenous variables, which are functions of four exogenous variables.
- While useful as a textbook exercise, implementing this model does not illuminate the challenges faced when the `forecast` tools are used for a professional forecasting model of the US macroeconomy.
- I describe our motivation to implement such a model in work I have been doing with a major US financial institution, Citizens Financial Group (CFG).

- To gain perspective on the problem: the Stata manual description of `forecast` illustrates its application to the venerable Klein Model I, a structure described in many econometrics textbooks.
- Klein Model I contains a total of three estimated equations and four identities: seven endogenous variables, which are functions of four exogenous variables.
- While useful as a textbook exercise, implementing this model does not illuminate the challenges faced when the `forecast` tools are used for a professional forecasting model of the US macroeconomy.
- I describe our motivation to implement such a model in work I have been doing with a major US financial institution, Citizens Financial Group (CFG).

# Modeling the macroeconomy for CCAR scenarios

- In the aftermath of the financial crisis, the Dodd–Frank legislation imposed new requirements on the Federal Reserve System's scrutiny of the largest financial institutions.
- The Fed has implemented the "CCAR" (Comprehensive Capital Analysis and Review) process for the 30+ largest financial institutions, involving "stress tests" of their modeled performance under various scenarios for macroeconomic variables that will affect their balance sheet and income statement.
- Financial institutions participating in the CCAR review must model macroeconomic outcomes under the Fed's detailed scenarios based on a fully-specified macro model.

# Modeling the macroeconomy for CCAR scenarios

- In the aftermath of the financial crisis, the Dodd–Frank legislation imposed new requirements on the Federal Reserve System's scrutiny of the largest financial institutions.
- The Fed has implemented the "CCAR" (Comprehensive Capital Analysis and Review) process for the 30+ largest financial institutions, involving "stress tests" of their modeled performance under various scenarios for macroeconomic variables that will affect their balance sheet and income statement.
- Financial institutions participating in the CCAR review must model macroeconomic outcomes under the Fed's detailed scenarios based on a fully-specified macro model.

# Modeling the macroeconomy for CCAR scenarios

- In the aftermath of the financial crisis, the Dodd–Frank legislation imposed new requirements on the Federal Reserve System's scrutiny of the largest financial institutions.
- The Fed has implemented the "CCAR" (Comprehensive Capital Analysis and Review) process for the 30+ largest financial institutions, involving "stress tests" of their modeled performance under various scenarios for macroeconomic variables that will affect their balance sheet and income statement.
- Financial institutions participating in the CCAR review must model macroeconomic outcomes under the Fed's detailed scenarios based on a fully-specified macro model.

- CFG now produces macroeconomic forecasts and performs scenario analysis with an extended version of Prof. Ray Fair's highly respected model of the US macroeconomy.
- The Fair US Model produces internally consistent forecasts of many of the variables needed in CCAR analysis.
- We have extended the Fair US Model to incorporate additional variables that are required for modeling the CCAR scenarios. These variables are driven by key elements of the Fair US Model.

- CFG now produces macroeconomic forecasts and performs scenario analysis with an extended version of Prof. Ray Fair's highly respected model of the US macroeconomy.
- The Fair US Model produces internally consistent forecasts of many of the variables needed in CCAR analysis.
- We have extended the Fair US Model to incorporate additional variables that are required for modeling the CCAR scenarios. These variables are driven by key elements of the Fair US Model.

- CFG now produces macroeconomic forecasts and performs scenario analysis with an extended version of Prof. Ray Fair's highly respected model of the US macroeconomy.
- The Fair US Model produces internally consistent forecasts of many of the variables needed in CCAR analysis.
- We have extended the Fair US Model to incorporate additional variables that are required for modeling the CCAR scenarios. These variables are driven by key elements of the Fair US Model.

- The Fair US Model is a medium-size model of the macroeconomy, with about 30 estimated equations and a number of identities.
- The model captures the behavior of six sectors of the economy: households, nonfinancial firms, financial firms, the federal government, state and local governments and the foreign sector.
- The model explicitly considers the flow of funds among these sectors and balance sheet constraints, linking the US National Income and Product Accounts (NIPA) with the US Flow of Funds Accounts (FOF).
- The Fair US Model includes a number of stock variables in addition to the macroeconomic and sectoral flows.

- The Fair US Model is a medium-size model of the macroeconomy, with about 30 estimated equations and a number of identities.
- The model captures the behavior of six sectors of the economy: households, nonfinancial firms, financial firms, the federal government, state and local governments and the foreign sector.
- The model explicitly considers the flow of funds among these sectors and balance sheet constraints, linking the US National Income and Product Accounts (NIPA) with the US Flow of Funds Accounts (FOF).
- The Fair US Model includes a number of stock variables in addition to the macroeconomic and sectoral flows.

- The Fair US Model is a medium-size model of the macroeconomy, with about 30 estimated equations and a number of identities.
- The model captures the behavior of six sectors of the economy: households, nonfinancial firms, financial firms, the federal government, state and local governments and the foreign sector.
- The model explicitly considers the flow of funds among these sectors and balance sheet constraints, linking the US National Income and Product Accounts (NIPA) with the US Flow of Funds Accounts (FOF).
- The Fair US Model includes a number of stock variables in addition to the macroeconomic and sectoral flows.

- The Fair US Model is a medium-size model of the macroeconomy, with about 30 estimated equations and a number of identities.
- The model captures the behavior of six sectors of the economy: households, nonfinancial firms, financial firms, the federal government, state and local governments and the foreign sector.
- The model explicitly considers the flow of funds among these sectors and balance sheet constraints, linking the US National Income and Product Accounts (NIPA) with the US Flow of Funds Accounts (FOF).
- The Fair US Model includes a number of stock variables in addition to the macroeconomic and sectoral flows.

- CFG have implemented the Fair US Model using Stata's `forecast` suite, based on the 2014Q4 version of the model (see http://fairmodel.econ.yale.edu/mmm1.htm).
- Using the original Fair US Model and dataset, we can compute unconstrained dynamic forecasts for the 2015–2017 period that match those computed by his freely available FORTRAN code.

- CFG have implemented the Fair US Model using Stata's `forecast` suite, based on the 2014Q4 version of the model (see http://fairmodel.econ.yale.edu/mmm1.htm).
- Using the original Fair US Model and dataset, we can compute unconstrained dynamic forecasts for the 2015–2017 period that match those computed by his freely available FORTRAN code.

- Although the Fair US Model contains a rich set of macroeconomic and sectoral variables, it does not include all of the variables specified in the Fed's CCAR scenarios.
- Additional variables are needed for risk management of several lines of business.
- Accordingly, CFG have enhanced the Fair US Model with a number of additional stochastic equations and identities.
- These equations are driven by variables present in the original Fair US Model, implying that unconstrained forecasts of original variables will not be altered by CFG's additional equations.

- Although the Fair US Model contains a rich set of macroeconomic and sectoral variables, it does not include all of the variables specified in the Fed's CCAR scenarios.
- Additional variables are needed for risk management of several lines of business.
- Accordingly, CFG have enhanced the Fair US Model with a number of additional stochastic equations and identities.
- These equations are driven by variables present in the original Fair US Model, implying that unconstrained forecasts of original variables will not be altered by CFG's additional equations.

- Although the Fair US Model contains a rich set of macroeconomic and sectoral variables, it does not include all of the variables specified in the Fed's CCAR scenarios.
- Additional variables are needed for risk management of several lines of business.
- Accordingly, CFG have enhanced the Fair US Model with a number of additional stochastic equations and identities.
- These equations are driven by variables present in the original Fair US Model, implying that unconstrained forecasts of original variables will not be altered by CFG's additional equations.

- Although the Fair US Model contains a rich set of macroeconomic and sectoral variables, it does not include all of the variables specified in the Fed's CCAR scenarios.
- Additional variables are needed for risk management of several lines of business.
- Accordingly, CFG have enhanced the Fair US Model with a number of additional stochastic equations and identities.
- These equations are driven by variables present in the original Fair US Model, implying that unconstrained forecasts of original variables will not be altered by CFG's additional equations.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

To gain perspective on how this *medium-sized* econometric model differs from the textbook Klein Model I (3 equations, 7 endogenous variables, 4 exogenous variables), consider that our enhanced version of the Fair US model contains:

- 35 estimation results from `ivregress, regress, arima`
- plus 265 identities
- for a total of 300 endogenous variables
- as functions of 148 exogenous variables

We now describe how the model is employed in CFG's modeling efforts.

- To conduct scenario analysis under CCAR, the model's solutions for CCAR variables must be overridden by the Fed's specified trajectories.
- This implies that all other variables of the Fed's model must adjust to the trajectories of the scenario.

- To conduct scenario analysis under CCAR, the model's solutions for CCAR variables must be overridden by the Fed's specified trajectories.
- This implies that all other variables of the Fed's model must adjust to the trajectories of the scenario.

- For instance, the altered path of real GDP growth under the Fed's Baseline, Adverse and Severe Adverse scenarios will cause all components of GDP to follow different paths during the forecast period than in the original model's dynamic forecasts.
- This is very easily implemented by the `actuals` option of `forecast solve`, which causes any non-missing values of endogenous variables to be held fixed in the solution process.
- Our first attempt to impose CCAR scenarios was not successful, as we tried to use `forecast adjust` to specify paths for endogenous variables which are defined by identities. Recent updates to the documentation state that this cannot be done: only estimated endogenous variables may be modified by `forecast adjust`.

- For instance, the altered path of real GDP growth under the Fed's Baseline, Adverse and Severe Adverse scenarios will cause all components of GDP to follow different paths during the forecast period than in the original model's dynamic forecasts.
- This is very easily implemented by the `actuals` option of `forecast solve`, which causes any non-missing values of endogenous variables to be held fixed in the solution process.
- Our first attempt to impose CCAR scenarios was not successful, as we tried to use `forecast adjust` to specify paths for endogenous variables which are defined by identities. Recent updates to the documentation state that this cannot be done: only estimated endogenous variables may be modified by `forecast adjust`.

- For instance, the altered path of real GDP growth under the Fed's Baseline, Adverse and Severe Adverse scenarios will cause all components of GDP to follow different paths during the forecast period than in the original model's dynamic forecasts.
- This is very easily implemented by the `actuals` option of `forecast solve`, which causes any non-missing values of endogenous variables to be held fixed in the solution process.
- Our first attempt to impose CCAR scenarios was not successful, as we tried to use `forecast adjust` to specify paths for endogenous variables which are defined by identities. Recent updates to the documentation state that this cannot be done: only estimated endogenous variables may be modified by `forecast adjust`.

# In-sample backtests

- Analysis of the individual CFG custom equations (for brevity, not listed here) shows that they fit well over the estimation sample. We now consider their performance after integration with the full Fair US Model using a *backtest*.
- We construct a dynamic simulation of the model's 300 variables for the last 28 quarters of the estimation sample: 2008Q1–2014Q4.
- The model's fitted values over the simulation period are based on the model's own prior values, rather than the observed historical values.

# In-sample backtests

- Analysis of the individual CFG custom equations (for brevity, not listed here) shows that they fit well over the estimation sample. We now consider their performance after integration with the full Fair US Model using a *backtest*.
- We construct a dynamic simulation of the model's 300 variables for the last 28 quarters of the estimation sample: 2008Q1–2014Q4.
- The model's fitted values over the simulation period are based on the model's own prior values, rather than the observed historical values.

# In-sample backtests

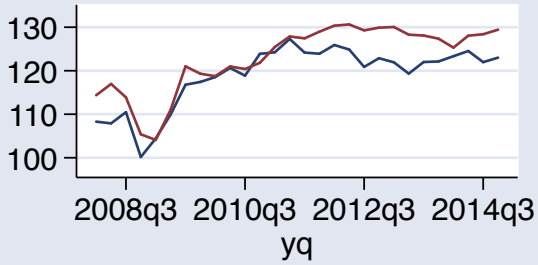
- Analysis of the individual CFG custom equations (for brevity, not listed here) shows that they fit well over the estimation sample. We now consider their performance after integration with the full Fair US Model using a *backtest*.
- We construct a dynamic simulation of the model's 300 variables for the last 28 quarters of the estimation sample: 2008Q1–2014Q4.
- The model's fitted values over the simulation period are based on the model's own prior values, rather than the observed historical values.

# Backtest: CFG Variables and Dynamic Forecasts

blue: actual, red: forecast)

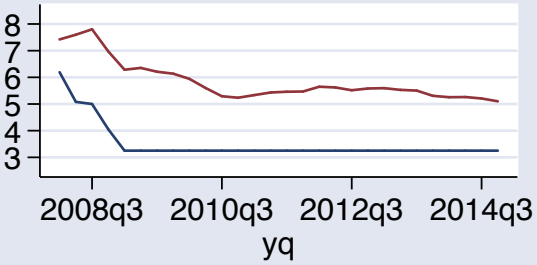
ALTSALES


MANHEIM

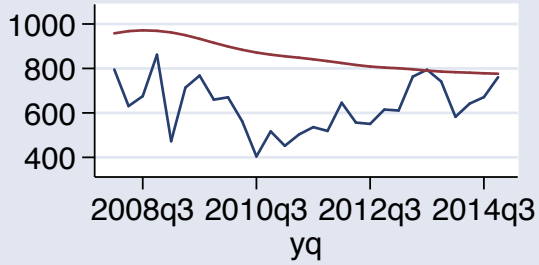



CREI

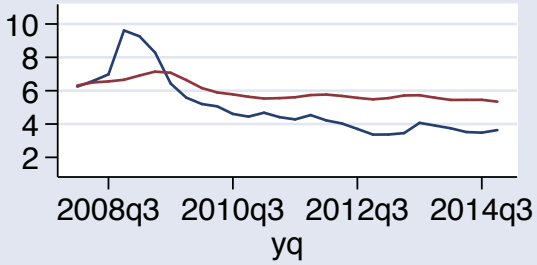

FRPRIME



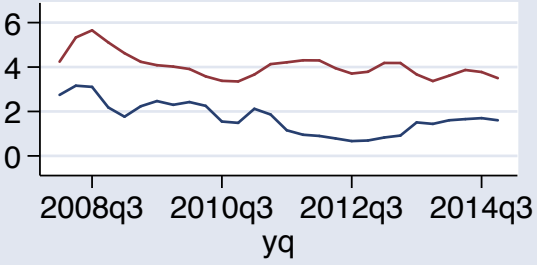

MORTORIG




RBBB




GS5




GS10


CORPPFT

- By starting the backtest in 2008Q1, we do not condition on the historical trajectories of macroeconomic variables during the financial crisis.
- Accordingly, the model does not capture the sharp drops in auto sales (ALTSALES) and mortgage originations (MORTORIG) caused by consumer credit constraints, nor the actual trajectory of interest rates (FRPRIME, RBBB, GS5, GS10) following unprecedented Fed actions.
- Despite the very poor performance of interest rate forecasts, the model does a reasonable job of fitting the Manheim used-car index, CREI (corporate real estate), and corporate profits.

- By starting the backtest in 2008Q1, we do not condition on the historical trajectories of macroeconomic variables during the financial crisis.
- Accordingly, the model does not capture the sharp drops in auto sales (ALTSALES) and mortgage originations (MORTORIG) caused by consumer credit constraints, nor the actual trajectory of interest rates (FRPRIME, RBBB, GS5, GS10) following unprecedented Fed actions.
- Despite the very poor performance of interest rate forecasts, the model does a reasonable job of fitting the Manheim used-car index, CREI (corporate real estate), and corporate profits.

- By starting the backtest in 2008Q1, we do not condition on the historical trajectories of macroeconomic variables during the financial crisis.
- Accordingly, the model does not capture the sharp drops in auto sales (ALTSALES) and mortgage originations (MORTORIG) caused by consumer credit constraints, nor the actual trajectory of interest rates (FRPRIME, RBBB, GS5, GS10) following unprecedented Fed actions.
- Despite the very poor performance of interest rate forecasts, the model does a reasonable job of fitting the Manheim used-car index, CREI (corporate real estate), and corporate profits.

# Computing unconstrained dynamic forecasts

- Before implementing CCAR scenarios, in which key variables are constrained to follow specified paths, we examine the out-of-sample performance of the model for the CFG custom variables.
- Dynamic forecasts over the 2015Q1–2017Q4 period produce trajectories for the model's 300 variables that are functions of exogenous variables' values and the model's logic.
- We implement this forecast using assumptions made in the published Fair US Model's forecasts for the next 12 quarters.

# Computing unconstrained dynamic forecasts

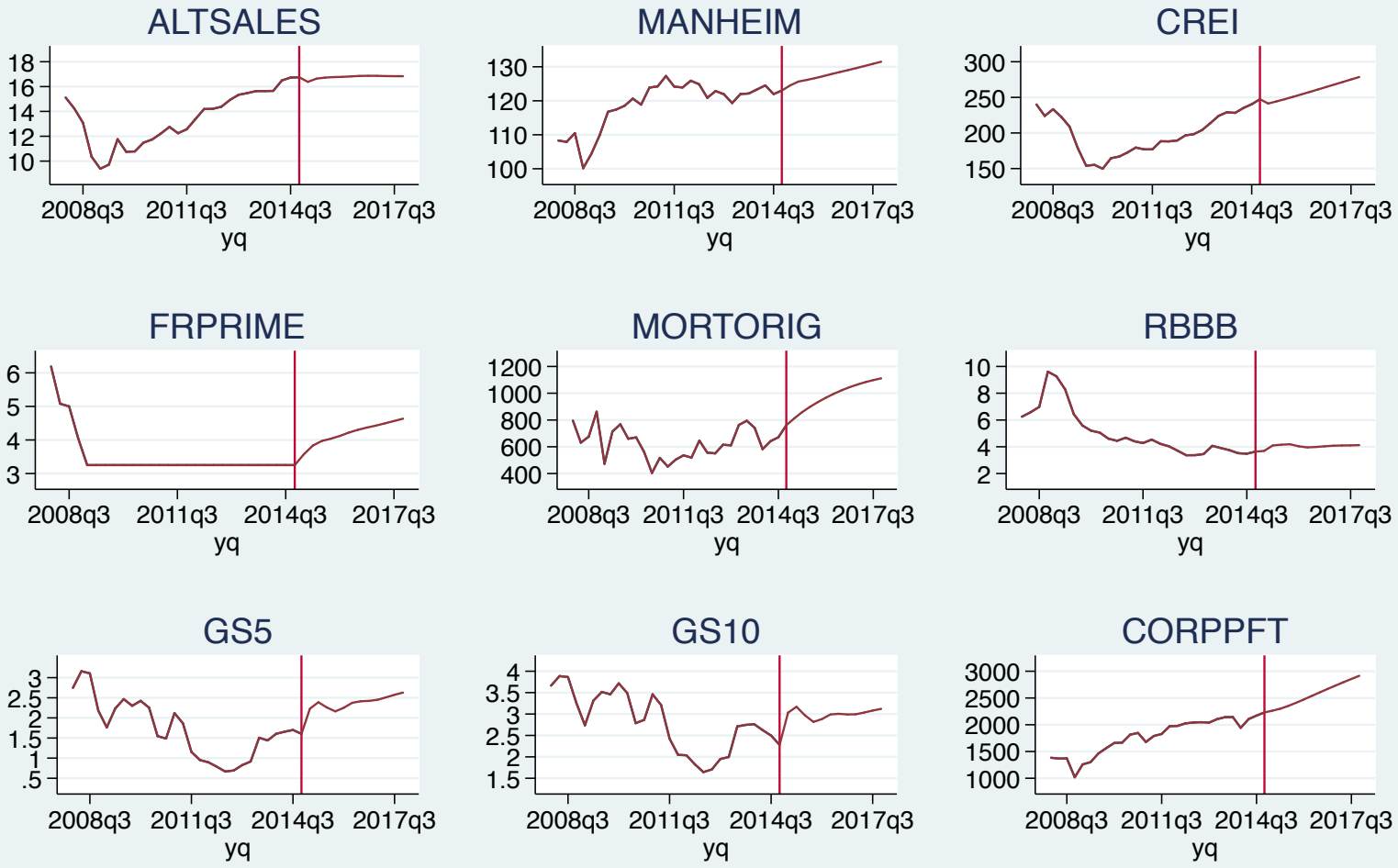
- Before implementing CCAR scenarios, in which key variables are constrained to follow specified paths, we examine the out-of-sample performance of the model for the CFG custom variables.
- Dynamic forecasts over the 2015Q1–2017Q4 period produce trajectories for the model's 300 variables that are functions of exogenous variables' values and the model's logic.
- We implement this forecast using assumptions made in the published Fair US Model's forecasts for the next 12 quarters.

# Computing unconstrained dynamic forecasts

- Before implementing CCAR scenarios, in which key variables are constrained to follow specified paths, we examine the out-of-sample performance of the model for the CFG custom variables.
- Dynamic forecasts over the 2015Q1–2017Q4 period produce trajectories for the model's 300 variables that are functions of exogenous variables' values and the model's logic.
- We implement this forecast using assumptions made in the published Fair US Model's forecasts for the next 12 quarters.

# CFG Variables and Dynamic Forecasts

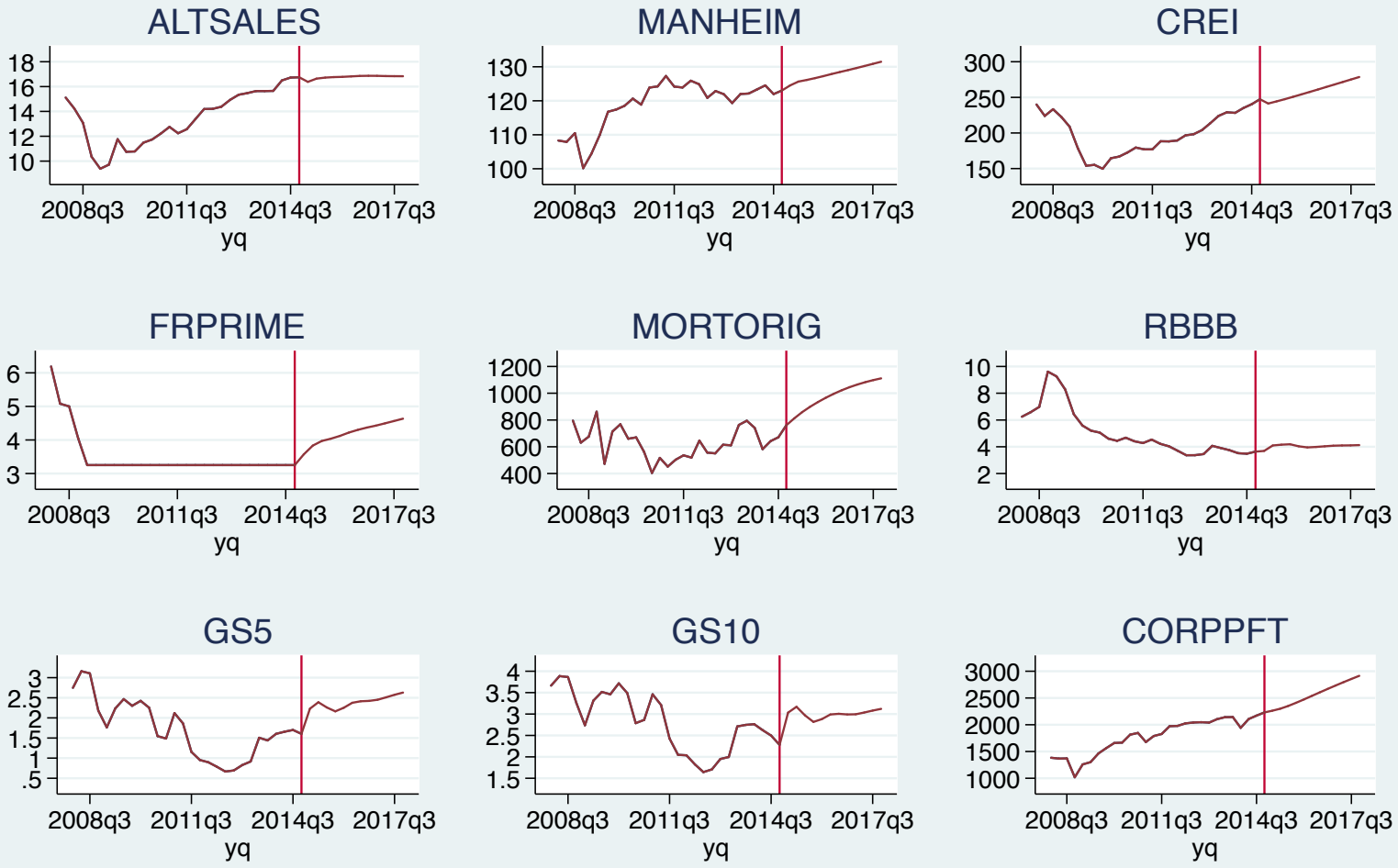
blue: actual, red: forecast)




All of the CFG custom variables' trajectories are sensible over the out-of-sample period.

# CFG Variables and Dynamic Forecasts

blue: actual, red: forecast)




All of the CFG custom variables' trajectories are sensible over the out-of-sample period.

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
  - Real GDP growth
  - Percentage change, GDP Deflator
  - Real Disposable Income growth
  - Unemployment rate
  - Commercial real estate price index
  - Housing price index
  - Bank prime rate
  - Corporate BBB bond rate
  - Five-year Treasury note rate
  - Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

- To implement the CCAR scenarios, we define the trajectories of several of the model's variables as those defined in each of the specified scenarios. These include:
- Real GDP growth
- Percentage change, GDP Deflator
- Real Disposable Income growth
- Unemployment rate
- Commercial real estate price index
- Housing price index
- Bank prime rate
- Corporate BBB bond rate
- Five-year Treasury note rate
- Ten-year Treasury note rate

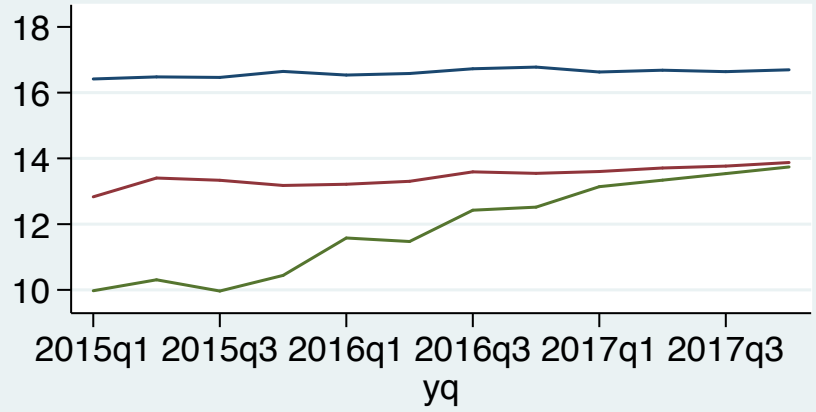
- Dynamic simulations are then performed for each of the CCAR scenarios: Baseline, Adverse, and Severe Adverse, over the forecast period 2015Q1–2017Q4.
- In these simulations, constraining the key variables force all other variables in the model to obey the NIPA, balance sheet and Flow-of-Funds constraints present in the Fair US Model's logic.
- To illustrate, we present the CFG custom variables that are not directly constrained by CCAR under the three scenarios.

- Dynamic simulations are then performed for each of the CCAR scenarios: Baseline, Adverse, and Severe Adverse, over the forecast period 2015Q1–2017Q4.
- In these simulations, constraining the key variables force all other variables in the model to obey the NIPA, balance sheet and Flow-of-Funds constraints present in the Fair US Model's logic.
- To illustrate, we present the CFG custom variables that are not directly constrained by CCAR under the three scenarios.

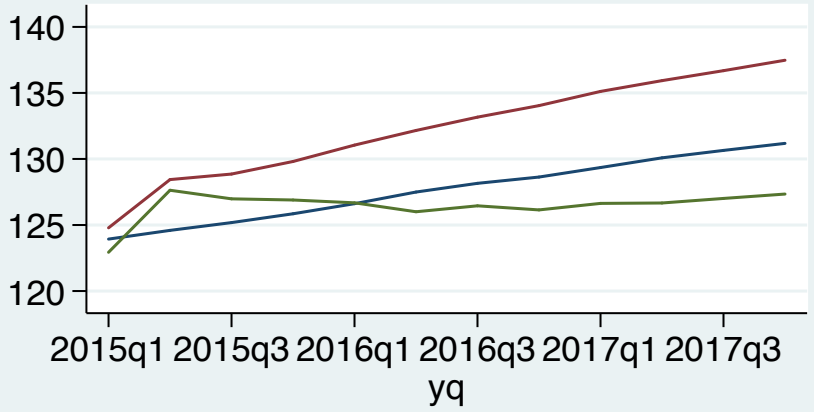
- Dynamic simulations are then performed for each of the CCAR scenarios: Baseline, Adverse, and Severe Adverse, over the forecast period 2015Q1–2017Q4.
- In these simulations, constraining the key variables force all other variables in the model to obey the NIPA, balance sheet and Flow-of-Funds constraints present in the Fair US Model's logic.
- To illustrate, we present the CFG custom variables that are not directly constrained by CCAR under the three scenarios.

# CFG Variables under CCAR Scenarios

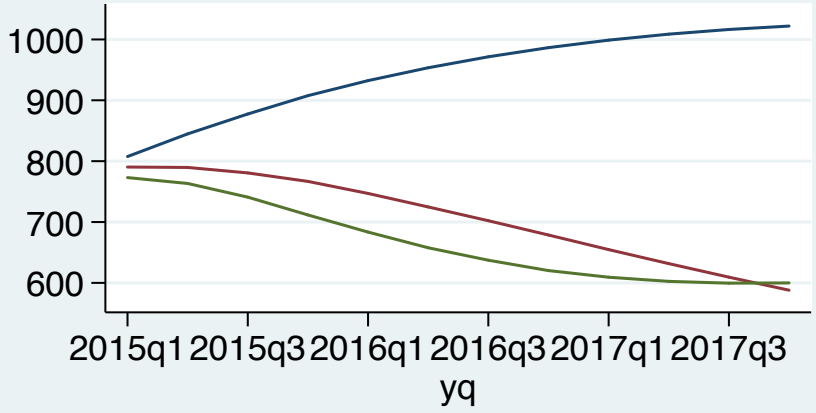
## Forecast: Light Vehicle Sales



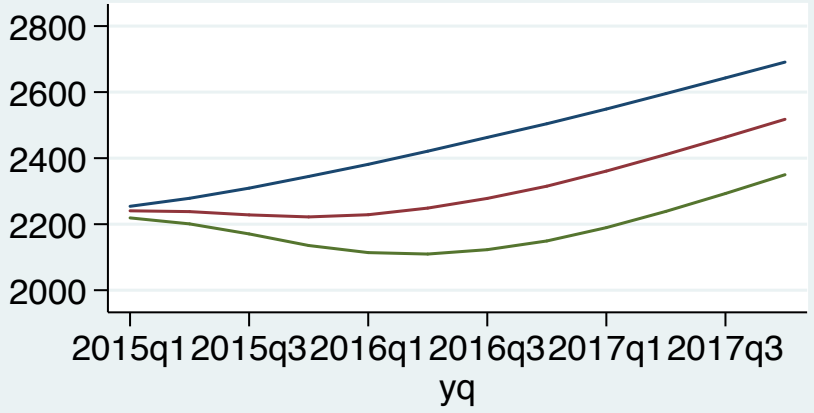

## Forecast: Used Car Price Index




## Forecast: Mortgage Originations




## Forecast: Corporate Profits




baseline adverse sev adverse

- The CFG custom variables show appropriate sensitivity to the trajectories of key macroeconomic factors across the three CCAR scenarios.
- This illustrates how the macroeconomic modeling capabilities required for CCAR purposes and general risk management may be developed in-house, reducing reliance on 'black box' third-party forecasts subject to regulators' critiques.

- The CFG custom variables show appropriate sensitivity to the trajectories of key macroeconomic factors across the three CCAR scenarios.
- This illustrates how the macroeconomic modeling capabilities required for CCAR purposes and general risk management may be developed in-house, reducing reliance on ‘black box’ third-party forecasts subject to regulators’ critiques.

# So far, so good?

I have illustrated our team's success in implementing the Fair US model, with enhancements, for CCAR scenario analysis. So what are the challenges?

# *So far, so good?*

I have illustrated our team's success in implementing the Fair US model, with enhancements, for CCAR scenario analysis. So what are the challenges?

- First, it took many hours of head-scratching and debugging to get to this point.
- As the Stata manual indicates, all exogenous variables must be non-missing for the periods in which `forecast solve` is to be used.
- Declaration of exogenous variables is optional, but strongly encouraged.
- However, if a variable that appears in an exogenous role in any equation is not declared as such and is missing, the model will not solve, and the resulting error specifies that dozens—or hundreds—of endogenous variables cannot be computed: it does not spell out where the missing values appear.

- First, it took many hours of head-scratching and debugging to get to this point.
- As the Stata manual indicates, all exogenous variables must be non-missing for the periods in which `forecast solve` is to be used.
- Declaration of exogenous variables is optional, but strongly encouraged.
- However, if a variable that appears in an exogenous role in any equation is not declared as such and is missing, the model will not solve, and the resulting error specifies that dozens—or hundreds—of endogenous variables cannot be computed: it does not spell out where the missing values appear.

- First, it took many hours of head-scratching and debugging to get to this point.
- As the Stata manual indicates, all exogenous variables must be non-missing for the periods in which `forecast solve` is to be used.
- Declaration of exogenous variables is optional, but strongly encouraged.
- However, if a variable that appears in an exogenous role in any equation is not declared as such and is missing, the model will not solve, and the resulting error specifies that dozens—or hundreds—of endogenous variables cannot be computed: it does not spell out where the missing values appear.

- First, it took many hours of head-scratching and debugging to get to this point.
- As the Stata manual indicates, all exogenous variables must be non-missing for the periods in which `forecast solve` is to be used.
- Declaration of exogenous variables is optional, but strongly encouraged.
- However, if a variable that appears in an exogenous role in any equation is not declared as such and is missing, the model will not solve, and the resulting error specifies that dozens—or hundreds—of endogenous variables cannot be computed: it does not spell out where the missing values appear.

- In a toy model like Klein Model I, finding such a problem is trivial. In a realistic sized model, it is exceedingly difficult to track down the missing values.
- Obviously Stata is aware of the structure of each equation and identity, and should be able to track down problems of missing values in any exogenous variable. At present, it does not do so.
- I would recommend that Stata's `forecast describe exogenous` should be enhanced to include all exogenous variables, classifying them as defined and undefined. This would greatly enhance users' ability to debug their model.
- Attempts to use `forecast solve` for periods in which there are missing values should clearly specify the *exogenous* variable(s) in which missing values appear, rather than indicating all the *endogenous* variables for which solutions cannot be produced.

- In a toy model like Klein Model I, finding such a problem is trivial. In a realistic sized model, it is exceedingly difficult to track down the missing values.
- Obviously Stata is aware of the structure of each equation and identity, and should be able to track down problems of missing values in any exogenous variable. At present, it does not do so.
- I would recommend that Stata's `forecast describe exogenous` should be enhanced to include all exogenous variables, classifying them as defined and undefined. This would greatly enhance users' ability to debug their model.
- Attempts to use `forecast solve` for periods in which there are missing values should clearly specify the *exogenous* variable(s) in which missing values appear, rather than indicating all the *endogenous* variables for which solutions cannot be produced.

- In a toy model like Klein Model I, finding such a problem is trivial. In a realistic sized model, it is exceedingly difficult to track down the missing values.
- Obviously Stata is aware of the structure of each equation and identity, and should be able to track down problems of missing values in any exogenous variable. At present, it does not do so.
- I would recommend that Stata's `forecast describe exogenous` should be enhanced to include all exogenous variables, classifying them as defined and undefined. This would greatly enhance users' ability to debug their model.
- Attempts to use `forecast solve` for periods in which there are missing values should clearly specify the *exogenous* variable(s) in which missing values appear, rather than indicating all the *endogenous* variables for which solutions cannot be produced.

- In a toy model like Klein Model I, finding such a problem is trivial. In a realistic sized model, it is exceedingly difficult to track down the missing values.
- Obviously Stata is aware of the structure of each equation and identity, and should be able to track down problems of missing values in any exogenous variable. At present, it does not do so.
- I would recommend that Stata's `forecast describe exogenous` should be enhanced to include all exogenous variables, classifying them as defined and undefined. This would greatly enhance users' ability to debug their model.
- Attempts to use `forecast solve` for periods in which there are missing values should clearly specify the *exogenous* variable(s) in which missing values appear, rather than indicating all the *endogenous* variables for which solutions cannot be produced.

- Second, in a model of this size, many similar variable names appear. Other econometric modeling software allows you to choose a variable and get a listing of what it depends upon (if endogenous) and what variables depend on it, highlighting everywhere in the model where this variable enters.
- This is a very useful feature, particularly if an endogenous variable is taking on unreasonable values in a dynamic simulation.
- As Stata is fully aware of how each endogenous variable enters the model, it should be possible to implement such a feature for a model that has been fully defined.
- I would recommend that such a feature be implemented to aid modelers' efforts.

- Second, in a model of this size, many similar variable names appear. Other econometric modeling software allows you to choose a variable and get a listing of what it depends upon (if endogenous) and what variables depend on it, highlighting everywhere in the model where this variable enters.
- This is a very useful feature, particularly if an endogenous variable is taking on unreasonable values in a dynamic simulation.
- As Stata is fully aware of how each endogenous variable enters the model, it should be possible to implement such a feature for a model that has been fully defined.
- I would recommend that such a feature be implemented to aid modelers' efforts.

- Second, in a model of this size, many similar variable names appear. Other econometric modeling software allows you to choose a variable and get a listing of what it depends upon (if endogenous) and what variables depend on it, highlighting everywhere in the model where this variable enters.
- This is a very useful feature, particularly if an endogenous variable is taking on unreasonable values in a dynamic simulation.
- As Stata is fully aware of how each endogenous variable enters the model, it should be possible to implement such a feature for a model that has been fully defined.
- I would recommend that such a feature be implemented to aid modelers' efforts.

- Second, in a model of this size, many similar variable names appear. Other econometric modeling software allows you to choose a variable and get a listing of what it depends upon (if endogenous) and what variables depend on it, highlighting everywhere in the model where this variable enters.
- This is a very useful feature, particularly if an endogenous variable is taking on unreasonable values in a dynamic simulation.
- As Stata is fully aware of how each endogenous variable enters the model, it should be possible to implement such a feature for a model that has been fully defined.
- I would recommend that such a feature be implemented to aid modelers' efforts.

- Third, one of the most frustrating aspects of `forecast` is that the model cannot be saved as an object and reloaded for further use.
- Although most of the information needed to rebuild the model can be saved to r-class macros by `forecast describe`, there is no command to save the model in a reusable form.
- The estimated equations can be saved to disk, as `.ster` files, so that reestimation is not required, but all of the `forecast...` commands defining the model must be given anew every time the model is to be used in a Stata session. These commands may be saved to a do-file with `forecast list`, but that do-file must be run to rebuild the model.
- Other econometric modeling software does not have this limitation.

- Third, one of the most frustrating aspects of `forecast` is that the model cannot be saved as an object and reloaded for further use.
- Although most of the information needed to rebuild the model can be saved to r-class macros by `forecast describe`, there is no command to save the model in a reusable form.
- The estimated equations can be saved to disk, as `.ster` files, so that reestimation is not required, but all of the `forecast...` commands defining the model must be given anew every time the model is to be used in a Stata session. These commands may be saved to a do-file with `forecast list`, but that do-file must be run to rebuild the model.
- Other econometric modeling software does not have this limitation.

- Third, one of the most frustrating aspects of `forecast` is that the model cannot be saved as an object and reloaded for further use.
- Although most of the information needed to rebuild the model can be saved to r-class macros by `forecast describe`, there is no command to save the model in a reusable form.
- The estimated equations can be saved to disk, as `.ster` files, so that reestimation is not required, but all of the `forecast...` commands defining the model must be given anew every time the model is to be used in a Stata session. These commands may be saved to a do-file with `forecast list`, but that do-file must be run to rebuild the model.
- Other econometric modeling software does not have this limitation.

- Third, one of the most frustrating aspects of `forecast` is that the model cannot be saved as an object and reloaded for further use.
- Although most of the information needed to rebuild the model can be saved to r-class macros by `forecast describe`, there is no command to save the model in a reusable form.
- The estimated equations can be saved to disk, as `.ster` files, so that reestimation is not required, but all of the `forecast...` commands defining the model must be given anew every time the model is to be used in a Stata session. These commands may be saved to a do-file with `forecast list`, but that do-file must be run to rebuild the model.
- Other econometric modeling software does not have this limitation.

- This limitation is out of line with other Stata commands, such as the ability to carry out a complex estimation and save its results in one session, and read the `.ster` file and compute `margins` in a second session.
- Likewise, `var` and `svar` commands allow one to save an `.irf` file containing impulse response functions, dynamic multipliers, and FEVDs, allowing production of tables and graphs without having to reestimate the original model.
- In our application, it was a bit tricky to modularize the modeling process by writing one do-file that estimated and defined the model, performed the unconstrained simulations, and a second do-file that performed scenario analysis.

- This limitation is out of line with other Stata commands, such as the ability to carry out a complex estimation and save its results in one session, and read the `.ster` file and compute `margins` in a second session.
- Likewise, `var` and `svar` commands allow one to save an `.irf` file containing impulse response functions, dynamic multipliers, and FEVDs, allowing production of tables and graphs without having to reestimate the original model.
- In our application, it was a bit tricky to modularize the modeling process by writing one do-file that estimated and defined the model, performed the unconstrained simulations, and a second do-file that performed scenario analysis.

- This limitation is out of line with other Stata commands, such as the ability to carry out a complex estimation and save its results in one session, and read the `.ster` file and compute `margins` in a second session.
- Likewise, `var` and `svar` commands allow one to save an `.irf` file containing impulse response functions, dynamic multipliers, and FEVDs, allowing production of tables and graphs without having to reestimate the original model.
- In our application, it was a bit tricky to modularize the modeling process by writing one do-file that estimated and defined the model, performed the unconstrained simulations, and a second do-file that performed scenario analysis.

- As warned in the manual, `clear all` removes the model's definition from memory, so that the second do-file must tread carefully to avoid disrupting the model in memory.
- It would be preferable to be able to merely load the predefined model (`forecast use`?) rather than having to define the entire structure each time the model is to be solved.
- I would recommend that this facility be added to the `forecast` suite.

- As warned in the manual, `clear all` removes the model's definition from memory, so that the second do-file must tread carefully to avoid disrupting the model in memory.
- It would be preferable to be able to merely load the predefined model (`forecast use`?) rather than having to define the entire structure each time the model is to be solved.
- I would recommend that this facility be added to the `forecast` suite.

- As warned in the manual, `clear all` removes the model's definition from memory, so that the second do-file must tread carefully to avoid disrupting the model in memory.
- It would be preferable to be able to merely load the predefined model (`forecast use`?) rather than having to define the entire structure each time the model is to be solved.
- I would recommend that this facility be added to the `forecast` suite.

Despite these challenges and grumbles, Stata's forecast suite is a very useful and elegant addition to Stata's tools for econometric modeling. Its advantages include:

- *SPEED:* solving the enhanced Fair model, 300 variables for each of 28 periods, takes only 21 seconds on a reasonably powerful (2.3 Ghz) laptop.
- Although not relevant for this application, the `forecast` suite handles forecasts of panel data (cross-country, cross-industry, etc.)
- The suite also implements stochastic simulation, with the capability to handle both additive and multiplicative (parameter) uncertainty.

Despite these challenges and grumbles, Stata's forecast suite is a very useful and elegant addition to Stata's tools for econometric modeling. Its advantages include:

- *SPEED:* solving the enhanced Fair model, 300 variables for each of 28 periods, takes only 21 seconds on a reasonably powerful (2.3 Ghz) laptop.
- Although not relevant for this application, the `forecast` suite handles forecasts of panel data (cross-country, cross-industry, etc.)
- The suite also implements stochastic simulation, with the capability to handle both additive and multiplicative (parameter) uncertainty.

Despite these challenges and grumbles, Stata's forecast suite is a very useful and elegant addition to Stata's tools for econometric modeling. Its advantages include:

- *SPEED:* solving the enhanced Fair model, 300 variables for each of 28 periods, takes only 21 seconds on a reasonably powerful (2.3 Ghz) laptop.
- Although not relevant for this application, the `forecast` suite handles forecasts of panel data (cross-country, cross-industry, etc.)
- The suite also implements stochastic simulation, with the capability to handle both additive and multiplicative (parameter) uncertainty.

Despite these challenges and grumbles, Stata's forecast suite is a very useful and elegant addition to Stata's tools for econometric modeling. Its advantages include:

- *SPEED:* solving the enhanced Fair model, 300 variables for each of 28 periods, takes only 21 seconds on a reasonably powerful (2.3 Ghz) laptop.
- Although not relevant for this application, the `forecast` suite handles forecasts of panel data (cross-country, cross-industry, etc.)
- The suite also implements stochastic simulation, with the capability to handle both additive and multiplicative (parameter) uncertainty.

In conclusion, I am very pleased with StataCorp's efforts in adding econometric model forecasting to a rich suite of time series tools. A bank which makes heavy use of Stata for loan-level and vintage-level econometric analysis can now implement macro modeling in Stata as well.

And I no longer have to apologize for the lack of econometric modeling tools when presenting Stata's capabilities to macroeconomists at the IMF and central banks. Thanks, StataCorp developers!

In conclusion, I am very pleased with StataCorp's efforts in adding econometric model forecasting to a rich suite of time series tools. A bank which makes heavy use of Stata for loan-level and vintage-level econometric analysis can now implement macro modeling in Stata as well.

And I no longer have to apologize for the lack of econometric modeling tools when presenting Stata's capabilities to macroeconomists at the IMF and central banks. Thanks, StataCorp developers!